# VISION SYSTEMS FOR MOBILITY APPLICATIONS

A Dissertation
Presented to
The Academic Faculty

by

Yosuke Yajima

In Partial Fulfillment
of the Requirements for the Degree
Master of Science in the
School of Mechanical Engineering

Georgia Institute of Technology
May 2020

# VISION SYSTEMS FOR MOBILITY APPLICATIONS

Approved by:

Dr. Bert Bras, Advisor
School of Mechanical Engineering
*Georgia Institute of Technology*

Dr. Jun Ueda
School of Mechanical Engineering
*Georgia Institute of Technology*

Dr. Richard Simmons
Strategic Energy Institute
*Georgia Institute of Technology*

Date Approved: [April 21, 2020]

[To the students of the Georgia Institute of Technology]

# ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Bert Bras for his continuous support of my research and his guidance on my thesis since I joined his lab as an undergraduate research assistant. With his support, I was able to work on his robotics research, and this project made me pursue a career in robotics field.

I would like to thank Ford Motor Company who provided an opportunity and funding for this project especially Jacob Mathews and Anne Marie Graham-Hudak. I would also like to thank my teammate, Bryan Cochran, Jesus Pacheco, and Zulfiqar Zadi who worked with me to build the robot for this research project. Without your help and support, this robotics project cannot be accomplished. I would like to thank other lab mates in the Sustainable Design and Manufacturing Lab for their advice and feedback.

I would also like to thank Dr. Jun Ueda and Dr. Richard Simmons for serving as research committee members.

Finally, I would like to thank my family in Japan for their support. I would also like to thank my friends in Atlanta, RoboJackets, and GT Solar Car Racing Team that helped me gained invaluable and fun engineering experience.

# TABLE OF CONTENTS

**ACKNOWLEDGEMENTS** **iv**

**LIST OF TABLES** **vii**

**LIST OF FIGURES** **viii**

**SUMMARY** **ix**

**CHAPTER 1. Introduction** **1**
**1.1 Opportunities and Potentials in Automated Charging Stations** **1**
**1.2 Challenge in Computer Vision for Automated Charging Stations** **4**
**1.3 Robot Setup** **5**
1.3.1 Hardware and Software 5
1.3.2 Camera Calibration 9
**1.4 Proposed Solutions** **9**
**1.5 Thesis Organization** **9**

**CHAPTER 2. Literature Review** **11**
**2.1 Computer Vision in Robotics** **11**
2.1.1 Object Detection and Recognition 11
2.1.2 3-D Reconstruction for Determining Depth Distance 15
**2.2 Review of Industry and Research in the field of Autonomous Charging Stations** **18**
**2.3 Conclusion** **23**

**CHAPTER 3. Methods** **24**
**3.1 Vision System Overview** **24**
**3.2 Data Preparation** **25**
3.2.1 Types of Object Class used in Object Detection Module 26
3.2.2 Data for Object Detection and Classification 28
3.2.3 Data for Depth Estimation 29
**3.3 Charging Inlet Detection and Classification** **31**
3.3.1 System Overview 31
3.3.2 Description of You Only Look Once 32
3.3.3 Training Process and Pre-trained Weight 33
**3.4 Distance Estimation** **34**
3.4.1 Depth Estimation System Overview 34
3.4.2 Distance Estimation Using Triangle Similarity 35
3.4.3 How Artificial Neural Networks work 36
3.4.4 Distance Estimation using Neural Networks 38
**3.5 Conclusion** **39**

**CHAPTER 4. Results and discussion** **40**
**4.1 Object Detection and Recognition** **40**

4.1.1 Evaluation Method 40
4.1.2 Results for Circular and Squared Shaped Fender 43
**4.2 Depth Estimation 47**
4.2.1 Evaluation Method 47
4.2.2 Results 48
**4.3 Conclusion 53**

**CHAPTER 5. Conclusion and future work 55**
**5.1 Conclusion 55**
5.1.1 Object Detection and Recognition 55
5.1.2 Depth Reconstruction 56
**5.2 Future Work 56**

**APPENDIX 58**

**REFERENCES 66**

# LIST OF TABLES

Table 1 – Range of Popular electric vehicles in the United States [7] ..... 3
Table 2 – List of Hardware and Software used in the vision system..... 8
Table 3 - Results for Object Detection Model..... 44
Table 4 – Overall Results..... 46
Table 5 – Results for Depth Estimation Tests ..... 50
Table 6 – Result for Closed Squared Shaped Fender Class..... 58
Table 7 – Result for Closed Circular Shaped Fender Class..... 59
Table 8 - Results for Open Squared Shaped Fender Class ..... 59
Table 9 - Results for Open Circular Shaped Fenders ..... 60
Table 10 – Results for Fuel Door Released Object..... 60
Table 11 – Results for Socket Object ..... 61

# LIST OF FIGURES

Figure 1 – PHV Sales in the United States [3] ..... 1
Figure 2 – Number of EV Charging Stations in US [4] ..... 2
Figure 3 – High Level System Overview ..... 6
Figure 4 – Jetson TX2 [8] ..... 7
Figure 5 – Automated Charging Robot [14] ..... 8
Figure 6 – Autonomous Charging Stations developed by Stable [27] ..... 19
Figure 7 – e-smartConnect developed by Volkswagen AG [29] ..... 20
Figure 8 – Automated Charging Station [30] ..... 21
Figure 9 – A Snake like Charging Robot [32] ..... 22
Figure 10 – Underbody Autonomous Charging Station developed by Volterio [34] ..... 23
Figure 11 – High Level Vision System Software Architecture ..... 25
Figure 12 – An Example for Red Fender ..... 27
Figure 13 – An Example for Champagne Fender ..... 27
Figure 14 – Overview of Data ..... 29
Figure 15 - Example of Depth Measurement ..... 30
Figure 16 – High Level Control Flowchart for Charging Inlet Detection ..... 31
Figure 17 – YOLO Process Flowchart [35][34] ..... 33
Figure 18 – Darknet-53 model [36] ..... 34
Figure 19 – Depth Estimation System Overview ..... 35
Figure 20 – Pin Hole Camera Model [37] ..... 36
Figure 21 – Artificial Neural Networks ..... 38
Figure 22 – Two Layered Neural Network Model ..... 39
Figure 23 – Intersection of Union ..... 41
Figure 24 – Examples of True Positive, False Positive, and False Negative [38] ..... 41
Figure 25 – Example of the Precision and Recall Curve ..... 42
Figure 26 - Precision, Recall, Average Precision Comparison ..... 43
Figure 27 – An example of the blurred image ..... 45
Figure 28 – Depth Estimation Results ..... 48
Figure 29 – Examples of Different Lighting Conditions ..... 49
Figure 30 – Results of the Percent Errors ..... 50
Figure 31 - Depth Distributions ..... 52
Figure 32 – Results of Standard Deviations ..... 53
Figure 33 – An example of the closed circular red fender in the outdoor environment ... 62
Figure 34 – An example of the closed circular red fender in the indoor environment ..... 62
Figure 35 – An example of the closed circular red fender in the indoor environment ..... 63
Figure 36 – An example of the closed circular blue fender int he indoor environment ... 63
Figure 37 – An example of the closed circular blue fender in the outdoor environment . 64
Figure 38 – An example of the closed champagne fender in the outdoor environment ... 64
Figure 39 – An example of the opened champagne fender in the outdoor environment .. 65

# SUMMARY

Electric vehicles have been developed by many automakers and startup companies in recent years. These vehicles are often equipped with advanced driver assistance systems (ADAS) or automated driving systems to help improve driver safety from traffic accidents. The ADAS or automated driving systems have a potential to improve car safety because 94 to 96 % of car accidents were caused by human errors in the United States in 2016. [1] However, there are a few companies that design and manufacture electric charging stations for autonomous electric vehicles. With the rapid development of electric and driverless vehicles, there will be a high demand for developing automated charging infrastructure to meet customers' demand. This thesis proposes a new computer vision-based software solution to automate charging stations using object detection and depth estimation techniques.

The purpose of this thesis is to design a vision software system for hands-free robot-controlled charging stations. With a camera based controlled system, the vision system sends a location and a depth of a target charging point so that the charging robot can insert its charging plug into the charging point of an electrical vehicle. The vision software system consists of the object detection and the depth estimation systems. The object detection method identifies and localizes the target charging point. The depth estimation method estimates the distance between the charging robot and the charging point. The vision system can also identify different shapes and colors of charging points as well as the condition of the charging point such as open or close in indoor and outdoor environments.

Additionally, the vision system communicates with the robot to send the distance information between the charging point and the end effector of the charging robot.

Test results show that the object detection model has a precision score of 95% and a recall score of 78%. The depth estimation model has an average percent error of 5.48%. The dataset for validating each model is collected using a parallel robotic arm charging robot developed by students from the Sustainable Design and Manufacturing lab. The camera system for the object detection and the depth estimation modules are both installed on the parallel robotic arm charging robot. Using deep learning and computer vision techniques, the vision system provides a potential solution to automate the charging station with minimal sensors and the total cost of installation.

# CHAPTER 1. INTRODUCTION

## 1.1 Opportunities and Potentials in Automated Charging Stations

Plug-in electric vehicles (PEVs) sales have been reached to one million units in the United States in October 2018. [2] With a rapid development of PEVs,

Figure 1 depicts a high expected demand for electric charging stations in near future. Currently the United States has approximately 47,000 units of public and workplace charging stations to recharge PEVs as shown in Figure 2. These public charging stations require users to recharge their vehicles by hands. With the growth of self-driving technology and advanced driver assistance systems, public charging systems can be automated to provide a safe and customer-friendly charging infrastructure. Automated charging stations in the United States are not publicly available, and several companies are working to commercialize their products.

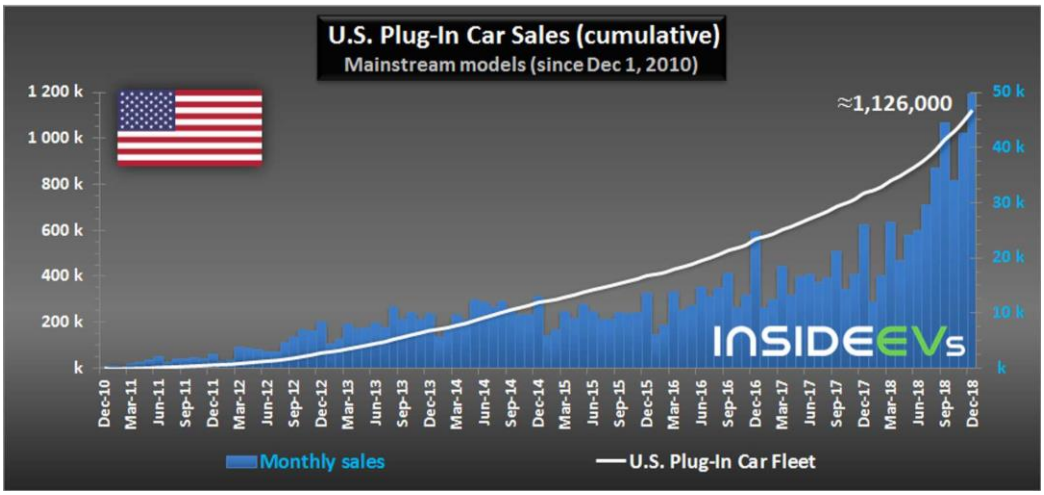


Figure 1 – PHV Sales in the United States [3]

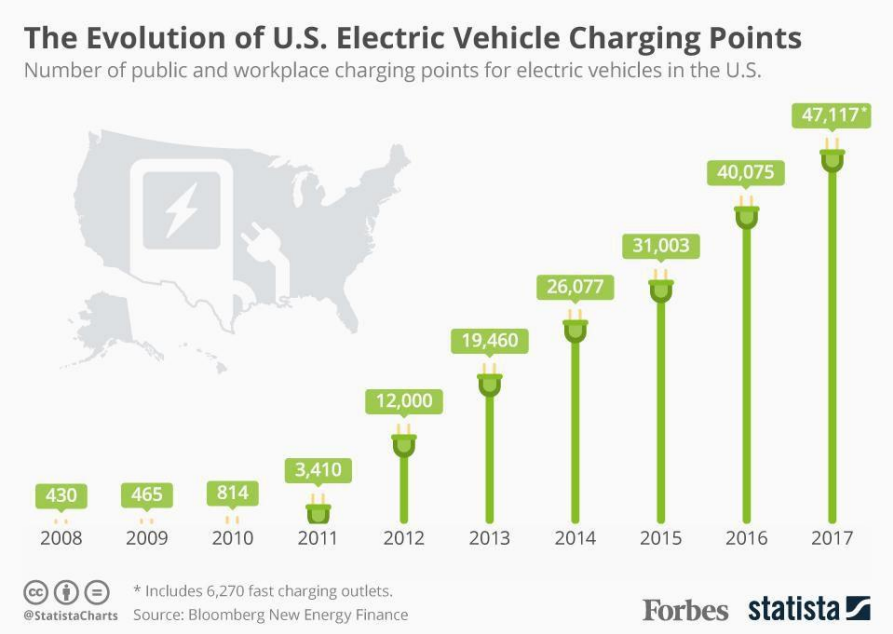


**Figure 2 – Number of EV Charging Stations in US [4]**

Beside the growth of the electric vehicle market, one advantage of automating the charging operation is to reduce a physical contact during the manual charging process. A list of driving ranges of popular electric vehicles is shown in Table 1. Because the electric vehicles sold in the United States have a limited driving ranges, electric vehicle owners are expected to charge their vehicle more frequently that gasoline-based vehicles. Automating the charging process allows the operator to avoid the risks of fire and electrical shock. With automated charging stations, the charging station can improve electrical safety.

In addition to electric safety, automated charging stations and autonomous vehicles together can provide a user-friendly and smooth recharging service. It is estimated that about 10 million autonomous vehicles will be on the road as early as 2020. [5] With the increasing number of self-driving cars on public roads, the development of autonomous parking systems with automated charging stations can create an efficient recharging cycle.

For example, Hyundai Motor Group announced their prototype of the automated valet parking system along with their charging system. [6] The automated parking system communicates with self-driving cars to identify the battery status, and the automated parking system navigates each vehicle to the appropriate parking spot. Because electric vehicle drivers often face overcrowding of charging locations, this system can optimize recharging operations by relocating the fully charged vehicle from the charging station to the other vacant parking spot to let other awaiting electric vehicles charge. When the customers need their vehicle, they can bring their car through the phone application. With several examples of benefits and needs for automated charging stations, there are potentials and opportunities of automated charging stations.

**Table 1 – Range of Popular electric vehicles in the United States [7]**

| Electric car model (base level) | Range (miles per charge) |
|---|---|
| Nissan Leaf | 150 miles |
| Chevrolet Bolt | 238 miles |
| Tesla Model S | 285 miles |
| Tesla Model X | 255 miles |

## 1.2 Challenge in Computer Vision for Automated Charging Stations

When it comes to building automated electric charging stations, it is necessary to implement a robust and accurate software solution to assist the charging process. This paper presents a low-cost perception-based robot charging station to recharge electric vehicles. The parallel robotic arm charging robot was originally designed by the Sustainable Design and Manufacturing lab, and two webcams are used along with the object detection and the depth estimation system. A list of the robot hardware and software is described in the robot setup section. The monocular camera sensors are used over other sensors (lidar, radar, ultrasonic sensor, etc.) because of target objects and limited space on the robot. The charging point contains information about colors and shapes. The camera sensor takes a picture that stores the color and texture information. This information is useful for identifying the location of the charging point. Regarding the space for the sensors, the current robot has limited space for mounting sensors. Because of this limitation, other sensors such as stereo vision cameras and lidar sensors cannot be installed to the end effector. Moving to the research questions, two major research challenge associated with the vision system are addressed below.

- Object detection and classification for charging points
- Depth estimation between the robot end effector and the target charging point

The object detection is a study of identifying and localizing a target object in an image. A bounding box is often placed over predicted target object. Numerous object detection approaches have been developed using traditional machine learning and modern neural networks approaches. Another major challenge in the vision system is a 3-D reconstruction

from a 2-D image. The depth reconstruction is a study of recovering a depth in an image. An image taken by a camera usually loses its depth dimension. Similar to the object detection, the recent development of deep learning provides numerous approaches to recover a missing depth information in an image.

## 1.3 Robot Setup

This section focuses on details of the automated charging robot and its hardware used as a part of the vision software system in this research. The important hardware and the software used in this research are presented. Camera calibration for camera sensors are performed using a chessboard, and its details are described in the camera calibration section.

### *1.3.1 Hardware and Software*

This section focuses on the hardware and the software used as a part of the vision software system. The overall system overview in Figure 3 shows a relationship between the vision software and the robot software. The vision software and the robot software communicate with each other through ROS messages that is a set of software library to build a robot software. The vision software has two main hardware: Jetson TX2 and two Logitech Webcams.

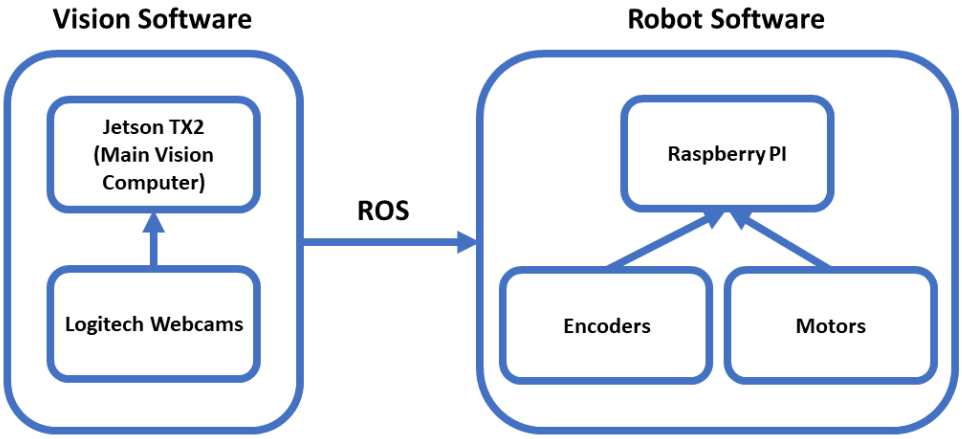


**Figure 3 – High Level System Overview**

The Jetson TX2 Developer Kit in Figure 4 is used as a primary computer to process the object detection and the depth estimation system during the robot operation. This computer contains a software package called JetPack 3.3, that comes with CUDA 9.0 as GPUs drivers, OpenCV3.3 as a computer vision library, and Ubuntu 16.04 as a Linux OS. The ROS is used as a primary robot communication tool to transmit information between camera sensors and the computer.

**Figure 4 – Jetson TX2 [8]**

Another important hardware in the vision system is the camera. Two Logitech C920 HD Pro Webcams are cameras sensors used in this automated charging project. One camera is attached to the top of the robot, and another camera is attached to the end of the robot end effector shown in Figure 5. A camera driver called cv_camera [9] is used to turn on the camera sensors. Regarding the object detection and the depth estimation software, the YOLOV3 and darknet [10] [11] are used to train the object detection model. The darknet_ros software packaged [12] is used to convert the object detection information to the ROS message so that the computer can understand information from the camera sensors. The labelImg [13] is used as a labeling tool to prepare annotations for training the object detector model. A summary of the hardware and the software used in the vision system is shown in Table 2.

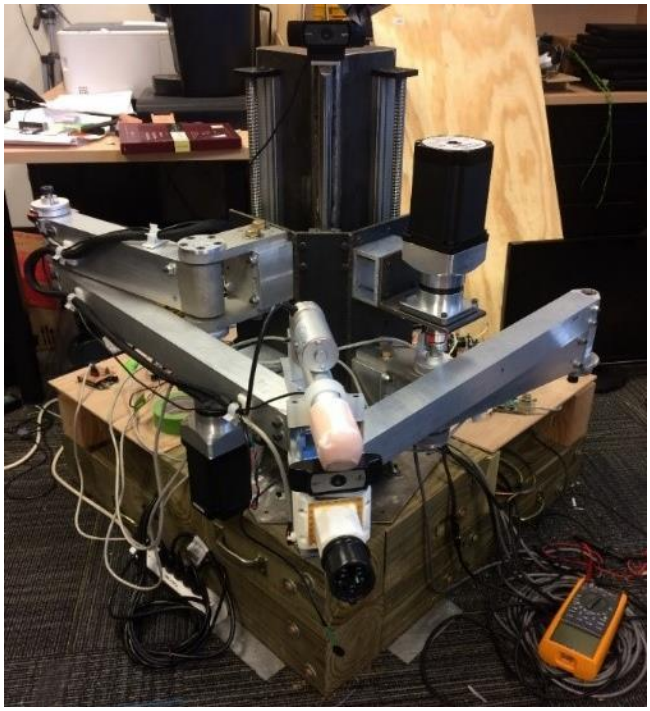

**Figure 5 – Automated Charging Robot [14]**

**Table 2 – List of Hardware and Software used in the vision system**

| Hardware | Software |
|---|---|
| TX2 Developer Kit | ROS Kinetics |
| Two Logitech camera C920 | Ubuntu 16.04 |
| | Python 2.7 |
| | YOLOV3 and darknet |
| | cv_camera |
| | darknet_ros |
| | labelImg |

### *1.3.2 Camera Calibration*

This section describes the camera calibration method to minimize a distortion on images produced by default camera sensors. A chessboard with 4-25mm squares-8x6 vertices and 9x7 squares [15] is used to calibrate the webcams. In addition to the camera calibration board, a software package called the camera_calibration is used [16]. This software package generates a camera matrix and a distortion matrix used to calibrate camera parameters.

## 1.4 Proposed Solutions

This paper presents a deep learning-based object detection using the YOLOV3 to identify the charging point of the vehicle. This method can identify the condition of the charging point as well as a fuel door release button and a charging socket of the electrical vehicle. Regarding the depth estimation, a bounding box of YOLOV3 is used to estimate the depth between the end effector of the charging robot to the target charging point. The object detection and the depth estimation are both used to assist the opening and closing sequence of the charging point. With the object detection and the depth estimation techniques, the charging robot can plug its charger into the target charging point.

## 1.5 Thesis Organization

This thesis presents a literature review of automated charging stations and computer vision in robotics, methods of the object detection and the depth estimation, and results and discussions of the object detection and the depth estimation with a custom dataset collected from robot experiments.

In the literature review, the review of traditional and modern object detection and depth estimation methods are discussed with examples. Current prototype and modern computer vision technology in automated charging stations from industry and research are also discussed. Following the literature review, the detail description of the object detection and the depth estimation algorithms are discussed in the method section. Results and discussions section show a performance of the object detection and the depth estimation system by testing the vision software with the automated parallel robot charging station. The conclusion and future work section describe a future work to improve some of the vision software system.

# CHAPTER 2. LITERATURE REVIEW

This section provides a summary of current research and previous study done on the topic of computer vision in robotics and automated charging robot systems. With a recent development of machine learning and deep learning technology, computer vision becomes a popular study applied in many robotics system and software applications. Object detection and depth estimation techniques are often used to create a map around the robot. The object detection identifies objects in the real world through an image or a video captured by camera sensors. The depth estimation is a study of reconstructing a 3-D environment from a 2-D image. Regarding the recent development of automated charging stations in industries and research labs, the camera sensors are primary used to detect and localize a position of charging point on a vehicle. The study of computer vision about the object detection and the depth reconstruction are presented first. The development of the automated charging stations in industries and research are presented next.

## 2.1 Computer Vision in Robotics

### *2.1.1 Object Detection and Recognition*

Object detection is a field of study in computer vision to classify and localize an object within an image that is captured with a camera. Due to the evolution in the area of deep learning and machine learning, many software systems use an application of object detection algorithms such as a face recognition and an autonomous vehicle. In the next paragraph, existing traditional and modern object detection models are discussed with examples such as Faster R-CNN and YOLO.

Kartik Umesh Sharma et al. published a review for the object detection system in an image. Their review focuses on a traditional approach to detect a target object in an image, and the literature review does not cover a modern object detection system using deep neural networks. There are four major object detection techniques presented in this review: sliding window-based object detection, Contour-based object detection, Graph-based object detection, and Context-based object detection. The sliding window-based object detection searches through the whole image to find the target object. Their result showed that the sliding window-based object detection required higher execution time due to searching the whole image. Because this method can handle one object at a time, detecting multiple objects requires a higher image processing time and more memory space is required. The contour-based object detection used a feature of contour from the images to search target objects. This method works well when the target object has a distinct shape and edges. The graph-based object detection uses a model that decomposes a target object into several parts and each part is represented by a graph vertex. The context-based object detection uses a model that contains key contexts and features of target object to detect the target objects. The review shows that the traditional approach to detect a target object often requires more computation time and memory space, and it may not be the best solution for mobile robot applications. [17]

Karanbir Chahal et al. published a review on modern object detections using deep learning technique. In their review, many different types of object detection systems based on the deep learning approach are presented. These object detection systems belong to either two-step detectors method or one step detectors method. The single-step method approach detects a target object and its location in a single step. The two-step approach divides the

single step approach into two steps. In the first step, the object detector searches a target object, and multiple regions of images are selected based on a higher probability of the target object in an image. Then the second step verifies if there is the target object presented in the image based on selected regions from the first step. The advantage of using the two-step approach is a higher accuracy and precision over the single step detectors. While the single step approach is better in terms of its accuracy, the single step approach is faster and more memory efficient resulting a better solution for a mobile robot with a limited computing power. Examples of the single step object detector are Single Shot Multi-Box Detector (SSD), You Only Look Once (YOLO), and Retina Net. Examples of the two-step object detectors are Region Convolutional Network (R-CNN), Fast RCNN, and Region Fully Convolutional Network (R-FCN). Based on their review, the object detection approaches using a deep neural networks are more suitable in mobile robot applications. [18]

Md. Hazrat Ali et al. conducted a research on developing a serial robot manipulator based on vision system which is attached at the robot gripper. This research focused on implementations of vision system to the existing sorting industrial robotic arms (Modern industrial robot Scorbot-ER 9 Pro) which can pick and place target objects. Vision systems are primary used as an object recognition and color detection to assist sorting the object using the webcam. The image processing techniques are applied to the raw image using the MATLAB's computer vision library such as converting images to binary images, calculating the area of object, and calculating the area ratio to identify the shape of target object out of circular, triangular and square shapes. The multiple USB cameras are implemented to the sorting robotic arm to reconstruct a depth distance from the images.

The main conclusion is that the vision system attached at the robot gripper assists a robotic arm to identify target object using MATLAB, USB cameras, basic MATLAB's image processing tools. [19]

Sungwoo Han proposed a low-cost advanced driver assistance system using USB webcam and lidar. He is a part of the EcoCAR Racing Team at Georgia Tech, building a semi-autonomous vehicle using 2016 Chevrolet Camaro and low-cost hardware such as NVIDIA Jetson TX2 and Logitech C920 Camera. A single camera is mainly used to develop his drive lane detection, vehicle detection, and distance estimation between a camera and preceding cars. In his vehicle detection system, You Only Look Once (YOLO), the object detection system based on the convolutional neural networks is used to precisely and accurately capture vehicles from the camera image. The YOLO creates a boundary box over the detected objects and tells what types of object is found. Additionally, his study includes a depth estimation method using the single web camera and the lidar. The depth between preceding vehicle and camera is calculated based on sensor fusion techniques. The main takeaway from his thesis is a use of the YOLO as object detection system and depth estimation. The YOLO has advantages over other object detection software in terms of fast processing speed and ease of integration to the low-cost hardware. The application of YOLO will be discussed more in depth in later section. [20]

Sandhya Sridhar developed an object detection and recognition software system for the vehicles detection and tracking system using machine learning and deep learning approach. He was a part of the EcoCAR3 team, and he was responsible for developing a vehicle detection and tracking system for the advanced driver assistance system. His thesis describes his training and validation strategies of the vehicle detection system, and he used

the Chevrolet Camaro from the EcoCAR3 team to conduct his experiment. He compared the method and result of vehicle detection system using the Cascade Object Detector (machine learning approach) and Deep Neural Network (DNN) Detector (deep learning approach) in his thesis. The cascade object detector uses a machine learning architecture developed by Viola Jones and the DNN Detector uses a deep learning architecture along with Max-Margin Object Detection (MMOD). He concluded that the DNN Detector performs better than the Cascade Object Detector in terms of precision and accuracy of placing the bounding box over the target objects. The Cascade Detector requires more fine tuning to optimize its performance while the DNN Detector required more processing time. In his future work section, he suggested to use alternative deep learning methods such as YOLO as an object detection tool. [21]

### *2.1.2 3-D Reconstruction for Determining Depth Distance*

Depth reconstruction from an image is another field of study in computer vision that is evolved with introduction to deep learning techniques. An image captured with a camera loses a dimension from 3-D world frame to 2-D image frame. There is a various technique to recover lost dimension from 2-D image. In the next paragraphs, a various method to reconstruct a depth dimension is discussed with examples.

Saloni Bahadur et al. published a review about various depth estimation methods for images using a features and cues of images. In order to recover the depth from 2-D images, their review shows two different features from images to recover the depth: stereo (triangulation) cues and monocular cues. Stereo cues are obtained from the stereo vision camera and it requires multiple images to recover the depth information. The monocular

cues do not require a stereo vision camera and it can be obtained from the single camera and the single image. Examples of the monocular cues includes texture, variations, gradients, defocus, and color. One of stereo cues-based methods is called the stereo matching. Two images taken by the stereo vision camera can be used to search the similar characteristics in images for estimating the depth. The example of monocular cues-based depth estimation is called object placement relation method. The object placement relation is a method that uses monocular cues to estimate the depth. This method is often used with a combination of machine learning and deep learning approaches. [22]

N. Jamwal et al. presented a review on depth estimation from 2-D images to 3-D images. They described that there are two types of depth estimation methods: monocular and multi-view method. The monocular method uses a single image captured by the camera to reconstruct the depth. The benefit of using a single monocular method is that the process time is less than the multi-view method. However, there is a downside of losing features and characteristics compared to the multi-view method. The Multi-view method uses more than two images to recover depth in images. This method is often used along with the stereo vision camera and it yields higher accuracy of reconstructing a depth in images. Their review discussed six types of depth estimation techniques to recover the depth from images. One method uses an analogy to human visual model that uses two cameras to estimate objects. This method requires cameras to be fixed at certain position and this may not be a good fit for mobile robot applications that requires a movement of cameras. The other interesting depth estimation strategy is to take a sequential picture from a single camera. The camera moves to a known distance and compare previous and current pictures to estimate a depth. Besides two examples above, another interesting method is to use an

auxiliary device such as lights to estimate the depth of images. Using a laser beam lights, the camera can capture some spots in the images that shows LED and using this points to estimate the depth. [23]

Mohammad A. Al-Jarrah proposed a software system based on a single monocular camera to construct 3-D environment using a small mobile robot. The software system calculates the distance between the camera and the target object as well as the angle between the camera and the target object. The distance and angle between the robot and target objects are calculated by taking two pictures and comparing the difference in vertices of objects. The first and second images are taken at different location. He concludes that a single monocular camera can estimate the depth and angles between the robot and target objects. [24]

Ashutosh Saxena et al. used the Markov Random Field (MRF) and the Laplacian model to estimates 3-D depth environment from single image using a deep learning approach. The MRF and the Laplacian model has an advantage in 3-D reconstruction problem because of insufficient local features and a lack of contextual information on the image. A supervised learning method was used to collect the depth map image from stereo vision camera as a ground truth image and the raw image from a single USB camera. His lab used a small scaled RC car to test the developed model at the unstructured outdoor environment. They also provided a detailed literature review for the depth estimation methods. Examples of important depth estimation techniques are described as following. Marco Quartulli et al. presented Bayesian information extraction method using a radar and laser to reconstruct a 3-D environment for the urban city map. D. Scharstein et al. presented a depth estimation

method using two or more images to recover the depth. Cornelis, Nico Leibe et al. showed a depth reconstruction method using a video sequence to model the 3-D city map. [25]

## 2.2 Review of Industry and Research in the field of Autonomous Charging Stations

Today, industry and academic institution corporate with each other to automate current charging stations for electric vehicles. Using a lidar and vision sensor, many research and industry came up a potential solution to automate charging cycle in the charging stations. In the next following section, a various proposed charging stations are discussed to illustrate existing charging stations.

Electrify America and Stable recently announced to deploy commercial automated charging stations for autonomous electric vehicles in San Francisco, CA in early 2020. The Electrify America is a subsidiary of Volkswagen Group of America and this company focuses on developing the infrastructure for electrical vehicles such as charging sites. The Stable is a start-up company for developing autonomous robotic charging stations using a 150kW DC fast charger. The DC fast charger is attached to the end effector of the robotic arms shown in Figure 6. The technical information about the charging method and sensors used in this robot is not described and not publicly available [26]


**Figure 6 – Autonomous Charging Stations developed by Stable [27]**

Volkswagen AG is developing an autonomous robot DC charging system for electric vehicles called e-smartConnect. The aim of this project is to provide an optimal recharging service for future electric vehicle consumers. The robotic manipulator called LBR iiwa from Kuka is used for this research, and the robotic arm has seven drive axles with torque sensors shown in Figure 7. The procedure to charge the electric vehicles are described as following: The electric vehicle communicated with the electric charging station to figure out where to park in conjunction with the automated parking system developed by the Volkswagen. Once the car is parked by the charging station, the robotics arm equipped with a camera mounted at the end effector grab the DC connector from the charging station. The robotic manipulator then inserts the DC connector to the socket of the vehicle. After the charging process is done, the robotic manipulator removes the DC connector and moves back to its original position. [28]


**Figure 7 – e-smartConnect developed by Volkswagen AG [29]**

Graz University of Technology, BMW AG, MAGNA Steyer Engineering, and KEBA AG developed and implement a prototype of the autonomous charging manipulator as a research project. In this prototype, the stereo vision camera and a monocular camera are used to detect the charging connectors on the test vehicle in Figure 8. The charging process starts with detecting the target vehicle. If the vehicle is detected, then the stereo vision camera finds the charging inlet on the vehicle and the end effector moves toward the charging inlet. Once the end effector reaches to the charging inlet, the camera attached at the end effector calculates the precise position of the charging inlet. The robot moves toward the charging inlet based on the end effector's camera. The end effector opens the charging door and the end effector inserts the charging connector into the charging socket. If the charging socket is fully connected to the charging connector, the charger starts transferring its electricity to the vehicle. When the charging operation completes, the end

effector closes the lid of the charging point and the end effector moves to its original position.


**Figure 8 – Automated Charging Station [30]**

Tesla Motors, the electric automaker from the United States, showed a prototype of the automated charging system. A snake like robotic arm charging robot in Figure 9 contains a charging connector at the gripper and this robotic arm moves toward the charging point. When the charging plug is connected to the charging socket, the charger starts to charge the vehicle. This product is not on the market and Tesla Motors does not announce a progress on the automated robotic arm charging station. The technical details and information about charging operation and sensors used in this prototype are not described by the Tesla Motors [31]

**Figure 9 – A Snake like Charging Robot [32]**

Volterio is developing an underbody autonomous charging station for electric vehicles as shown in Figure 10. Their goal is to design an affordable and a high-power charging system used at home. The charging connector is embedded at the gripper of the charging station, and this robotic manipulator can adjust its position and orientation for parking misalignment up to 0.5 by 0.5 m. The charging system communicates to the parked vehicle wirelessly to prepare for recharging operation. The charging operation starts with correcting the angle and position offset between the charging system and the vehicle socket. During this phase, the automated charging robot uses an ultrasound-based navigation system to navigate the end effector to the charging socket. When the final position of the end effector is confirmed, the automated charger starts recharging the vehicle. After the vehicle is recharged, the robot manipulator will move back to its original position. [33]


**Figure 10 – Underbody Autonomous Charging Station developed by Volterio [34]**

## 2.3 Conclusion

After reviewing object detection, depth estimation, and current development of the automated charging stations, there are various types of methods and technologies. In the field of the object detection and classification, the deep learning and machine learning approaches are widely used to classify object because of fast processing time and improved accuracy. Regarding the depth estimation, the stereo vision camera and single monocular camera are both used to reconstruct a depth in images. With recent performance of computers and GPUs, the deep learning approach is getting widely accepted in the field of mobile robotics. The development of automated charging stations shows various sensors and mechanical designs to recharge the vehicles. In the next section, the object detection using a deep learning approach and depth estimation algorithms are further discussed.

# CHAPTER 3. METHODS

Based on the literature review in the field of computer vision, robotics, and automated charging stations, the object detection and the depth estimation using a deep neural network demonstrate a possible solution to localize and classify the target charging inlet position in the 3-D coordinate system. In this chapter, a high-level overview of the vision system architecture is illustrated for the autonomous charging station. In the charging inlet detection section, details of the object detection system using the YOLOV3 is addressed to show how the object detection software identifies the charging point as well as other objects. Following the object detection method, the depth estimation section explains a three layers neural networks-based model to estimate the distance between the end effector and the target charging point. Details of data collection methods for both the charging inlet detection and the depth estimation are also discussed in this chapter.

## 3.1 Vision System Overview

The vision system package is composed of three modules: a camera module, an object detection module, and a depth estimation module shown in Figure 11. The camera module takes an input of an image to convert a preprocessed image so that the object detection module can use it to classify the target object in the image. The camera module uses the cv_camera package for the camera driver and the Logitech webcams for capturing an image. Following the camera modules, the object detection module used the YOLOV3 convolutional neural network (CNN) that identifies target objects such as charging points and other important objects on a vehicle fender. Examples of target objects include the condition of the charging port, the charging connector, and the fuel door release button. Details of targets object are further discussed in the data preparation section.

During the object detection module, the darknet_ros package is used to transfer the YOLOV3 detection outputs. These outputs contain information about detected objects with appropriate bounding box information. Once the object detection module outputs information about detected target objects, the centroid position for each detected target object is calculated in the captured image. Moving to the depth estimation module, with given bounding box parameters, the depth distance is estimated using the neural network that takes an input of the bounding box size. Finally, the depth and the centroid of target objects are transported to the robot for assisting the robot navigation. All modules communicate each other through the ROS. The next section describes the data preparation method, the object detection method, and the depth estimation method.

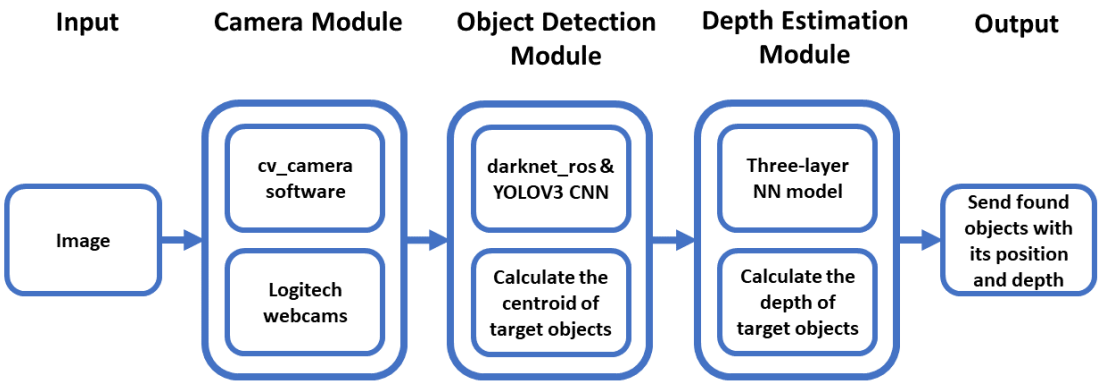


**Figure 11 – High Level Vision System Software Architecture**

## 3.2 Data Preparation

This section focuses on the data collection method for the object detection and the depth estimation module. To collect the data in the object detection module, three different types of car fenders are used in total. There are two circular shaped charging points with blue and red colored fenders and one champagne colored irregular shaped fender. Following the object detection module, the depth estimation module uses the red circular shaped car

fender alone. The data for the object detection and depth estimation are collected at indoor and outdoor environments. In the next two sections, the data collection method for object detection and the depth estimation method are addressed more in depth.

### *3.2.1 Types of Object Class used in Object Detection Module*

There are six major object categories used in the object detection module: open_circle, open_square, close_circle, close_square, fuel door release, and socket as shown in Figure 12 and Figure 13. Open and close object classes tell a condition of the charging port. The open class notifies that the robot completes a sequence of opening the fuel door, whereas the close class notifies that the robot completed a sequence of closing the fuel door. The bounding box of the open and close condition is used to calculate a centroid position of the charging point. The centroid position is used to navigate the end effector to the charging point on the vehicle. In addition to the condition of charging points, there are object classes for circle and square shaped fuel door. Moving to the fuel door release object class, it provides a centroid coordinate of fuel door release button that can be used to precisely locate the end effector to open the charging point. The centroid of the fuel door release provides more precise position when the end effector gets close to the car fender. Regarding the charging connector, the class object of the socket is used to finalize the position of end effector so that the charging connector can align and insert into the charging port. With these four class objects, the vision system can provide precise and accurate position of the charging point to recharge the electric vehicle.


**Figure 12 – An Example for Red Fender**

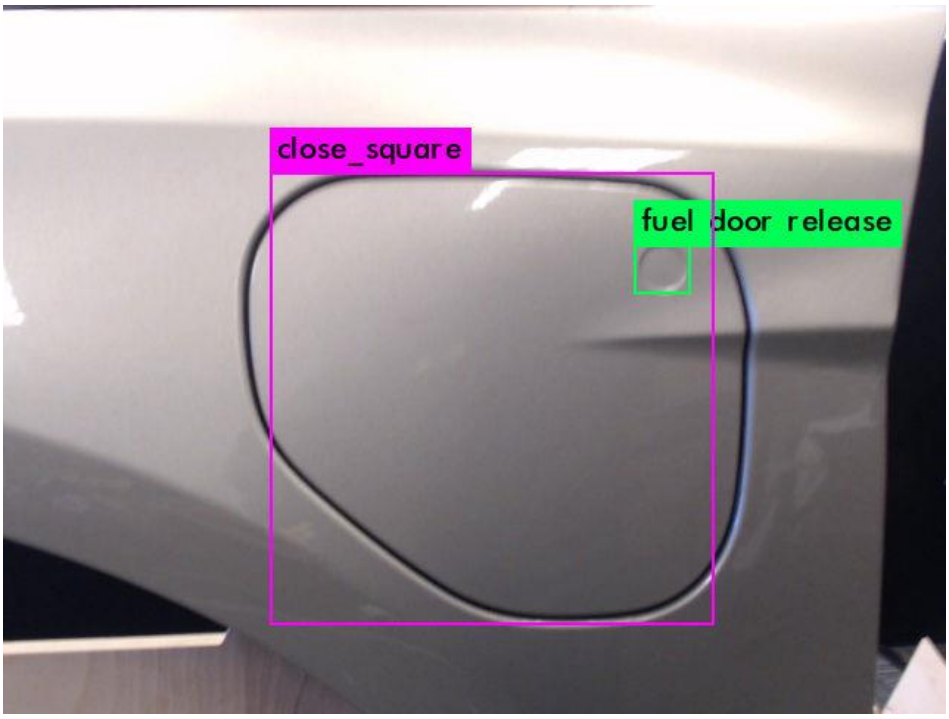


**Figure 13 – An Example for Champagne Fender**

### *3.2.2 Data for Object Detection and Classification*

The object detection relies on the YOLOV3 CNN which is a convolutional neural network model based on the supervised learning approach. The supervised learning approach uses a pair of input and annotated output data to train the model. In this case, input is a raw image that contains the target object such as the charging point, and the output is an annotated image that contains a bounding box over the target objects. The data consists of three major categories: there are background, open, and close shown in Figure 14. The background objects are considered as negative sample which does not contain target objects, whereas the open and close objects are considered as positive sample that contains target objects. The Background category includes background images such as parking lot without charging points. The benefit of including negative samples is to improve accuracy and precision of the convolutional neural network model. The open and close categories are considered as positive samples, and there are charging points, charging socket, and fuel door release. These categories have subcategories such as blue, red, champagne, circular, and square shaped fenders. Each category contains indoor with light turned on and off as well as outdoor environment with sunny and night conditions. In order to balance the data, the training data set contains same number of data categories. The data contains total of 4800 images. Each major category (background, open, close) contains 1600 images and each subcategory (indoor light, indoor dark, outdoor light, outdoor dark) contains 200 images.

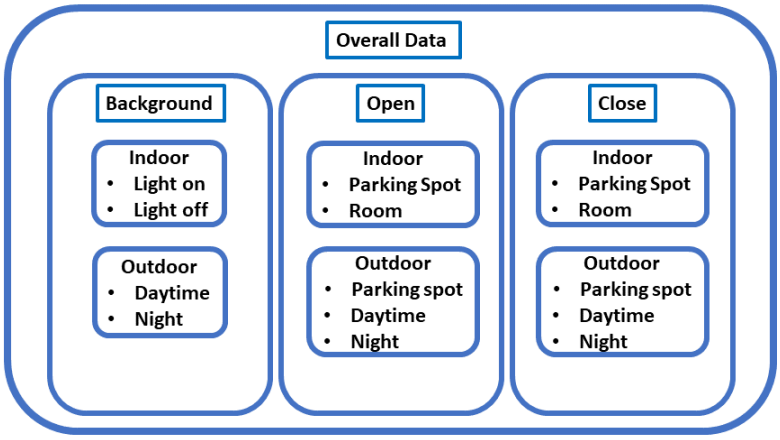


**Figure 14 – Overview of Data**

### *3.2.3 Data for Depth Estimation*

The data for the depth estimation module is collected with a red fender at both indoor and outdoor environment. The ground truth distance between the end effector and charging points are manually measured with a tape measurement tool in Figure 15. The instruction of data collection is described as follows.

1. The vehicle fender is placed in front of the charging point. The robot end effector is aligned along with the car charging points.
2. The robot then rises its height to align along with the middle of target charging point.
3. Following the robot and the fender setup, the depth between the charging point and the car fender are measured by every 1 inch from 20 inches to 7 inches away from the car fender. The length is then converted to metric unit in cm.

4. The width and height of the bounding box between the car fender and end effector is recorded every 1 inch. 25 sets of points are collected in the indoor and the outdoor environment, respectively.

These data are collected at the parking spot and lab environment. 40 sets of points are used for training the neural networks, and the 10 sets of points are used for validating neural networks performance.

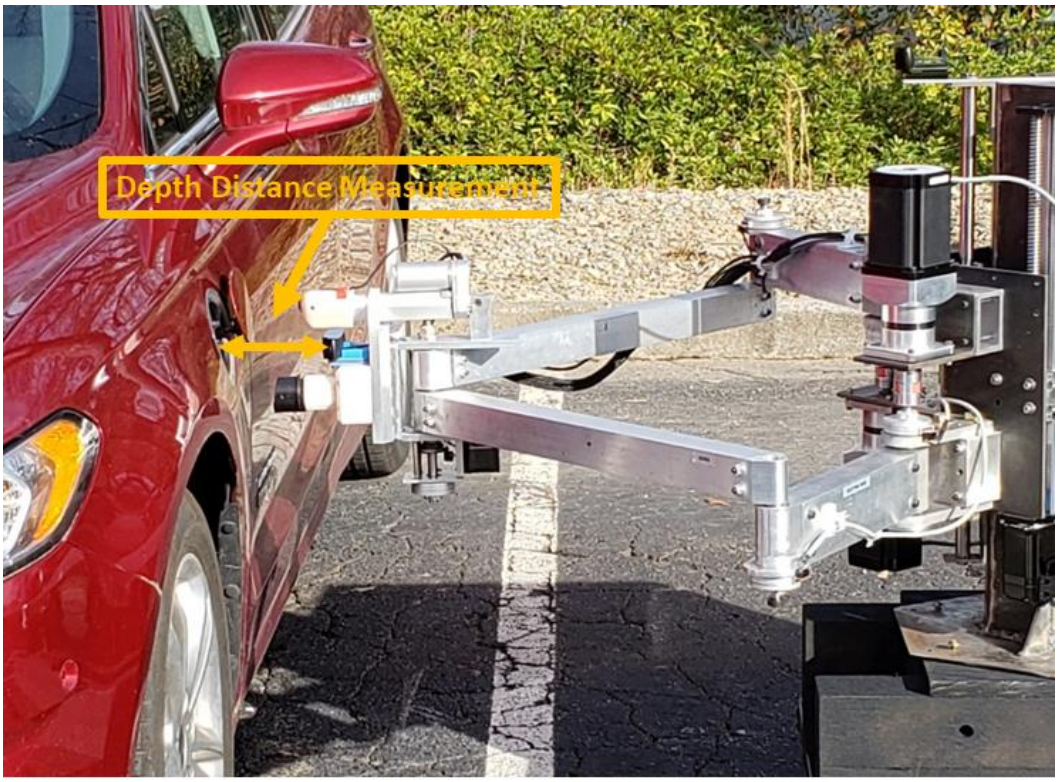


**Figure 15 - Example of Depth Measurement**

## 3.3 Charging Inlet Detection and Classification

### *3.3.1 System Overview*

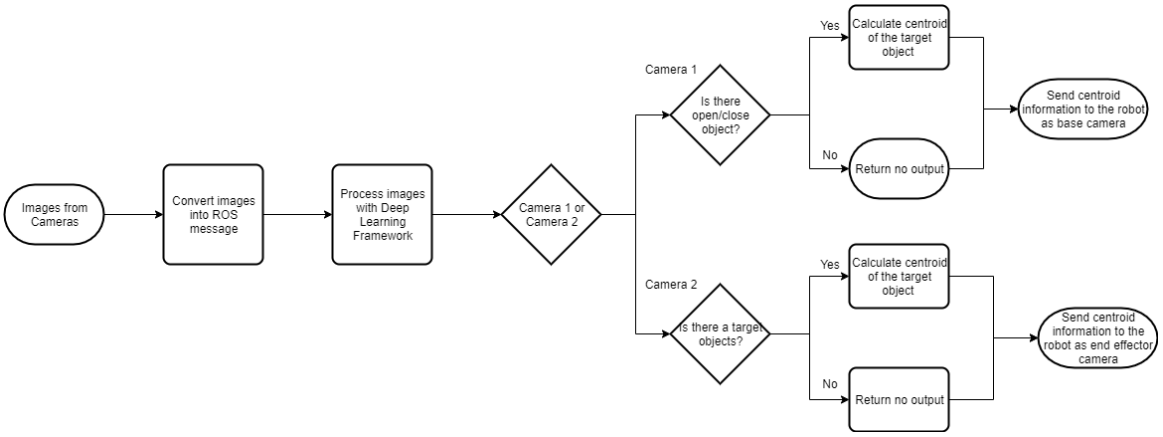


**Figure 16 – High Level Control Flowchart for Charging Inlet Detection**

This section addressed a control logic of the charging inlet detection system used in this project. The control flowchart for the object detection module in Figure 16 shows a step by step process. The first step is to obtain images from the first camera attached to the end effector and the second camera attached to the top of the charging robot. When the images are obtained, the images are converted to the ROS image message so the YOLOV3 model can use them as inputs. When the images are passed to the model, it outputs detected objects for each camera. If the images are taken by the first camera, then object detection system checks to see if there is a close or open class object. If there is a close or open object, the vision system calculates each centroid position based on the size of bounding box. If there is no target object, the vision system returns no centroid position. After this step is done, the vision system sends an information about overall detected targets with each corresponding centroid of target positions. Like the first camera, if the images are captured by the second camera, the vision system checks whether target objects such as a socket of charging inlet and a fuel door release detent are found in the detected images or not. The vision system calculates each centroid position of the target object while the vision system

returns to no centroid object if there is no target object in the images. Finally, the vision system sends all information about centroid of each detected target objects to the robot. In the next section, the detail and performance of the YOLOV3 is discussed.

### *3.3.2 Description of You Only Look Once*

The YOLO model is considered as a single shot object detector that takes an input image and pass it through the neural networks only once for predicting target objects in real-time. Unlike other object detection models such as sliding window approach, YOLO has an advantage of its processing speed to detect objects in real-time. YOLO model first takes and divides the input image into 13 by 13 regions. Each region predicts five bounding boxes with confidence scores and class predictions. There are total of 845 bounding boxes per input image. The confidence score tells a probability of the bounding box that encloses target objects, while the class prediction predicts a possible target object within the bounding box. After 845 bounding boxes are generated, the model applies a filter to eliminate bounding boxes that contain lower confidence score and prediction class. The non-max suppression is then applied to output one bounding box per each region. Using confidence score and class prediction, the model can classify and localize a bounding box over each predicted object in the image. This approach has an advantage over traditional CNN models based on sliding window techniques that requires multiple passes through the neural networks, because YOLO minimizes a processing time to classify and localize the target objects with corresponding bounding box especially for mobile robot applications.

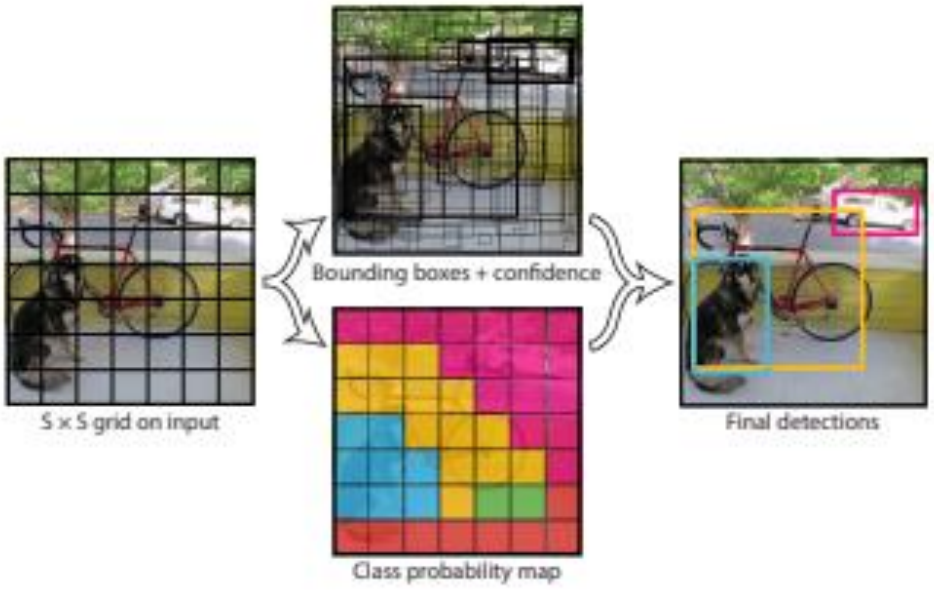


**Figure 17 – YOLO Process Flowchart [35][34]**

### *3.3.3 Training Process and Pre-trained Weight*

The YOLO pre-trained model named "darknet53. conv.74" and its machine learning software named Darknet were used to develop a custom object detection model. This model was proposed by Joseph Redmond, and its architecture is shown in Figure 18. This model can classify multiple objects and locate each detected object with a bounding box over the image. There are 53 convolutional layers and this model was trained from ImageNet dataset. In this thesis, the transfer learning approach is utilized to train a model that fits in the charging robot's tasks. Transfer learning is a technique to train a model from pre-trained model. Benefit of using pretrained model are to improve precision and reduce training time of the object detection system. The YOLO pre-trained weight was already trained with a large dataset with various objects. Using some of similar features in the convolutional neural networks, transfer learning can improve precision of the object detection as opposed to train the model with small dataset from scratch. Additionally, the transfer learning can take less time to train the model. In this thesis, I used my own data to fine tune the model.

| | Type | Filters | Size | Output |
|---|---|---|---|---|
| | Convolutional | 32 | 3 × 3 | 256 × 256 |
| | Convolutional | 64 | 3 × 3 / 2 | 128 × 128 |
| 1× | Convolutional | 32 | 1 × 1 | |
| | Convolutional | 64 | 3 × 3 | |
| | Residual | | | 128 × 128 |
| | Convolutional | 128 | 3 × 3 / 2 | 64 × 64 |
| 2× | Convolutional | 64 | 1 × 1 | |
| | Convolutional | 128 | 3 × 3 | |
| | Residual | | | 64 × 64 |
| | Convolutional | 256 | 3 × 3 / 2 | 32 × 32 |
| 8× | Convolutional | 128 | 1 × 1 | |
| | Convolutional | 256 | 3 × 3 | |
| | Residual | | | 32 × 32 |
| | Convolutional | 512 | 3 × 3 / 2 | 16 × 16 |
| 8× | Convolutional | 256 | 1 × 1 | |
| | Convolutional | 512 | 3 × 3 | |
| | Residual | | | 16 × 16 |
| | Convolutional | 1024 | 3 × 3 / 2 | 8 × 8 |
| 4× | Convolutional | 512 | 1 × 1 | |
| | Convolutional | 1024 | 3 × 3 | |
| | Residual | | | 8 × 8 |
| | Avgpool | | Global | |
| | Connected | | 1000 | |
| | Softmax | | | |

**Figure 18 – Darknet-53 model [36]**

*Part 2: Implementing object detection systems on the charging systems*

## 3.4 Distance Estimation

### *3.4.1 Depth Estimation System Overview*

Vision based distance measurement between the end effector and the target object is used for the robot navigation such as inserting the charging connector into the vehicle charging point. The depth estimation module estimates the distance between the charging point and the end effector equipped with a charging connector. The two-layer neural network is used to estimate the depth. The depth estimation module relies on the shape of the bounding box provided by the YOLO algorithms. Figure 19 shows an overview of the depth estimation module. When the object detection module provides detected objects with corresponding bounding box information, the depth estimation starts by checking if there is a close

condition of the charging points. The depth estimation module uses a close charging point to identify the distance between the end effector and charging points. If the charging point is classified as close, the width and height of the bounding box is passed to the neural network to predict the depth between the camera and charging point. The depth estimation module returns to no output if there is no close condition. After calculating the depth, the depth module checks if the depth belongs to the camera from the end effector or the base camera. The depth module finally sends depth information to the robot through the ROS communication tool.

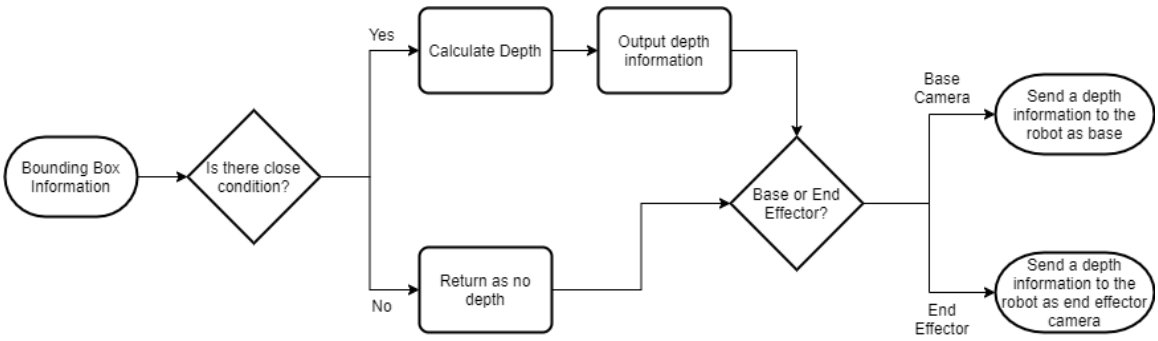


**Figure 19 – Depth Estimation System Overview**

### *3.4.2 Distance Estimation Using Triangle Similarity*

Using a pin hole camera model and triangle similarity, the monocular camera can approximate a distance between the robot and the target in an image. Figure 20 depicts the relationship between the world coordinate and camera coordinate system. The real width represents actual width of the charging point, while the pixel width represents bounding box size in an image. The focal length is a part of the intrinsic parameter, and it is calculated with the following equation.

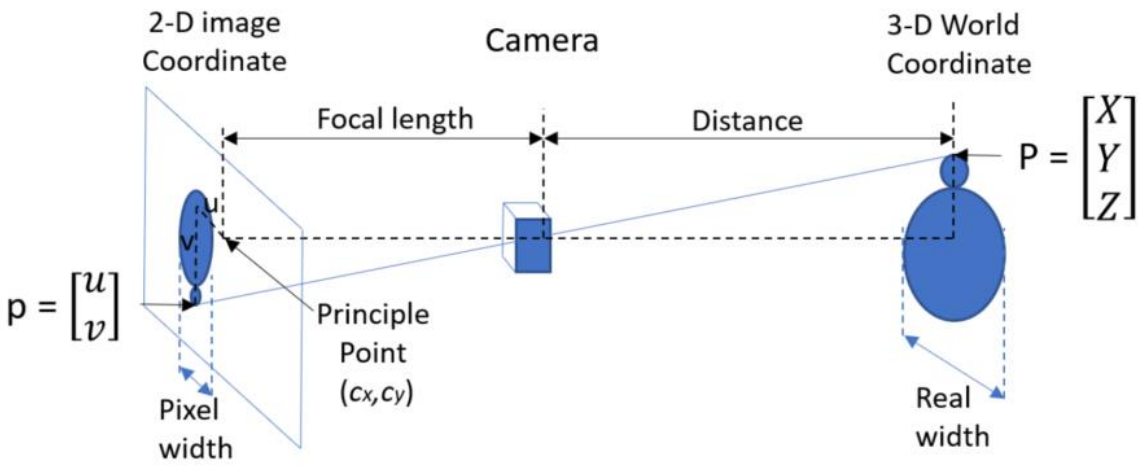


**Figure 20 – Pin Hole Camera Model [37]**

$$Focal\ Length = \frac{Bounding\ Box\ Width * Real\ Distance}{Real\ Width}$$

With known the focal length, bounding box width, and the real width, the distance is computed using the following equation. Real width corresponds to the actual width of charging points in meter. The bounding box width represents detected bounding boxes over the charging point in 2-D image coordinate system. Since the real width is fixed size and the bounding box width can be varied depend on the location of the camera, the equation can be used to approximate the distance between the end effector and the target object.

$$Distance\ between\ Robot\ and\ Target = \frac{Real\ Width * Focal\ Length}{Bounding\ Box\ Width\ (Pixel\ Width)}$$

### *3.4.3 How Artificial Neural Networks work*

Artificial neural network is one of machine learning techniques used in the deep learning field. Artificial neural network is a model inspired by the human brain and biological nervous system. Typically, the neural networks consist of an input layer, one or more

hidden layers, and one output layer shown in Figure 21. Each layer contains a set of nodes that interconnected with each other. The connections between each layer is often called as weight and biases that are parameters to be optimized during the training process. The weights and biases are used to calculate a weighted sum of previous nodes before next node process its calculation. The activate function takes a sum of calculated weights and biases and convert it to an output that can process to the next layers. This process repeats until the neural network reaches to the output layer. This whole process is also known as a forward propagation. After the forward propagation, the loss is computed using a cost function. Once the loss is calculated, the feedback process called the backpropagation updates each weights and biases to optimize its neural networks performance. The training process of the neural networks repeats a forward and backpropagation process until the model reaches to its best performance. In this thesis, trained model and its parameters are saved at every 1000 iterations. The next section introduces a distance estimation method using the artificial neural networks.

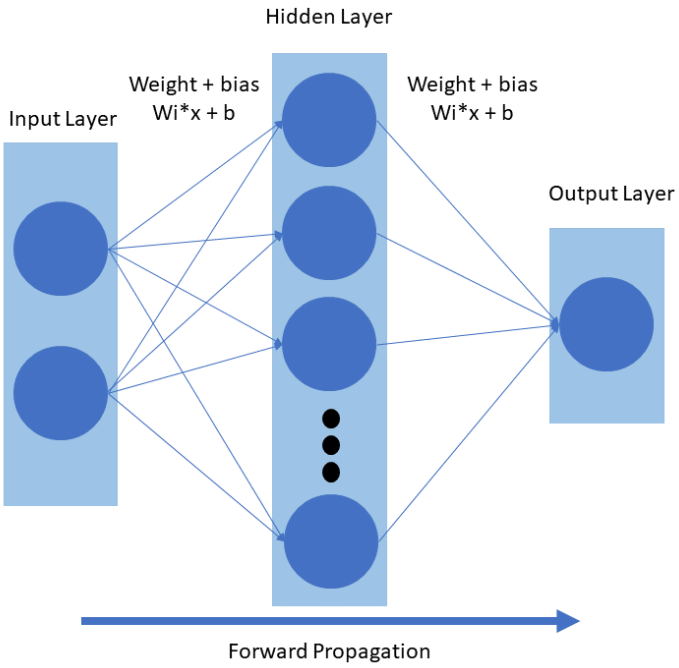


**Figure 21 – Artificial Neural Networks**

### *3.4.4 Distance Estimation using Neural Networks*

While the distance estimation using the triangle similarity shows an ability to estimate a depth using a monocular camera, the proposed neural network-based distance estimation predicts a distance without a focal length and a real width of the charging point. This neural network takes inputs of the width and the height of the bounding box to predict a distance between the robot and the target object. The Figure 22 shows a model architecture with two hidden layers that contain 100 nodes, respectively. Each hidden layer contains an activation function of the ReLU which converts a sum of weights and biases into an output that can be used for next layer. The loss is calculated using a mean squared error. The Adam optimization technique is used to update each weights and biases in this model. The output layer predicts a depth between the end effector of the robot to the charging point.

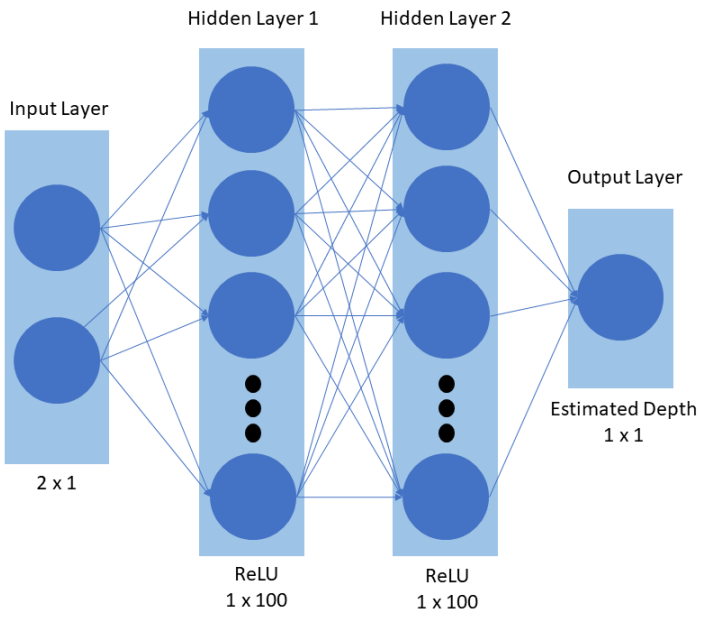


**Figure 22 – Two Layered Neural Network Model**

## 3.5 Conclusion

This chapter discusses methods to obtain the location of the target object using the YOLO algorithms and the artificial neural networks. In the next chapter, the validations and results of vision system are discussed with evaluation metrics.

# CHAPTER 4. RESULTS AND DISCUSSION

## 4.1 Object Detection and Recognition

### *4.1.1 Evaluation Method*

This section discusses a various evaluation metrics used to verify the trained object detection model. The evaluation method includes a mean average precision, a precision, and a recall. The precision metrics tells accuracy of the model prediction. The recall metric measures the performance of the model to find out all possible objects in the ground truth labels. The precision and recall metrics are both used to validate the performance of the neural networks. Precision and recall metrics are computed in the following equations.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

$$Recall = \frac{True\ Positibe}{True\ Positive + False\ Negative}$$

The true positive is defined as a correct prediction by the model, whereas the false positive is defined as an incorrect prediction by the model. The false negative is a case where the model fails to detect a target object in an image. Intersection of Union (IoU) in Figure 23 is an evaluation metric used to judge the true positive and false positive. The area of intersection is an overlapping area between a ground truth and a predicted bounding box, whereas the area of union is all area between a ground truth and a predicted bounding box. The IoU threshold is set to 0.5 in this experiment, and the true positive image is considered

if the IoU is larger than 0.5. Figure 24 contains several examples of the true positive, false positive, and false negative used in the precision and recall calculation.

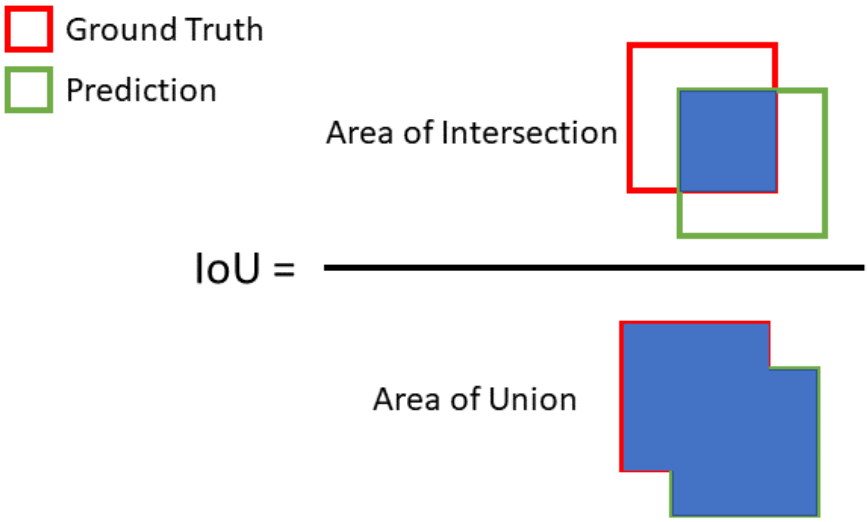


**Figure 23 – Intersection of Union**

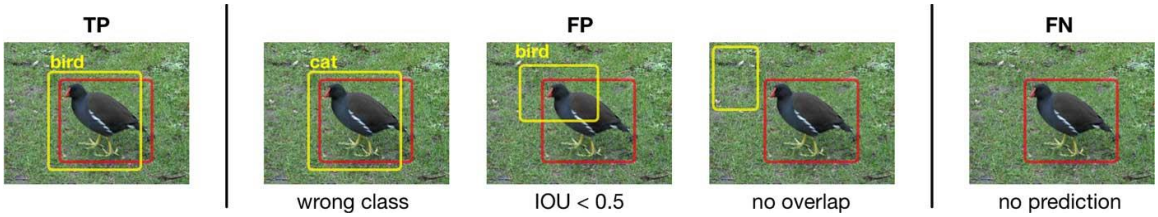


**Figure 24 – Examples of True Positive, False Positive, and False Negative [38]**

With precision and recall metrics, F-1 score is used to validate trained model at different iterations. During the training process, multiple models are saved at different iterations, and it is important to select a best model based on an optimal combination of precision and recall metrics. Since precision and recall metrics are dependent on each other, there is always a tradeoff between them. The F-1 score allows users to measure the best neural networks model in the equation below.

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Mean average precision (mAP) is used for alternative evaluation metric to the precision and recall metrics. To obtain the mean average precision, the precision and recall curve is used to calculate the average precision for each class such as charging points, a socket, and a fuel door release. The precision and recall curve contain data points of the precision and recall at different confidence threshold between 0 to 1 as shown in Figure 25.

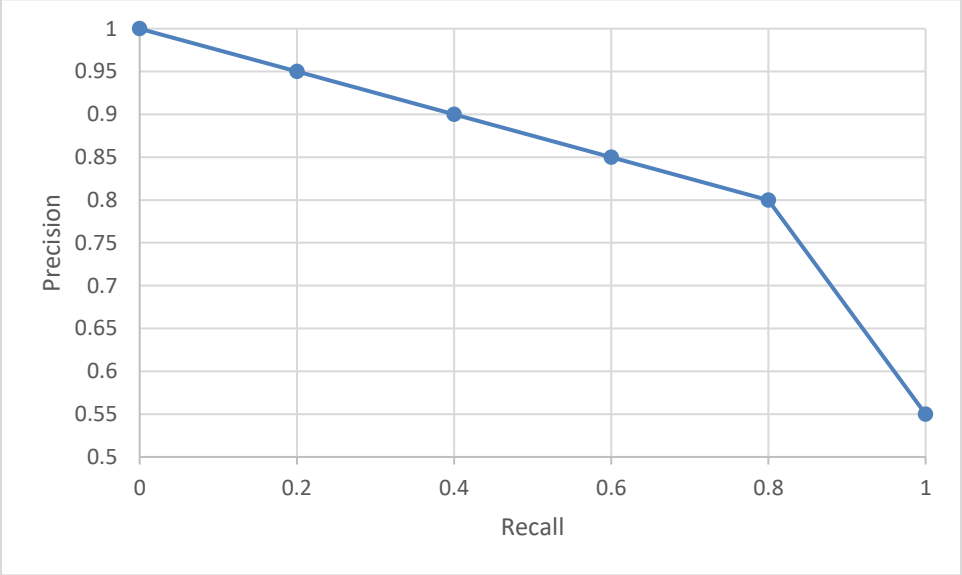


**Figure 25 – Example of the Precision and Recall Curve**

The average precision is also known as the area under the precision and recall curve. When the average precisions are calculated, the mean average precision is calculated as follows.

$$mAP = \frac{\sum_{q=1}^{Q} Average\ Precision}{Q}, Q = total\ object\ class$$

The mAP is defined as a sum of each average precision divided by the overall object class. This metric has an advantage over the precision and recall metric, because it considers all

object classes instead of individual object class. Since the model contains six different object classes, mean average precision is a suitable metric to evaluate the model. The target goal is to obtain minimum scores of 90 % for the precision, the recall, and the mean average precision score. In the next section, results are discussed with different evaluation metrics.

### *4.1.2 Results for Circular and Squared Shaped Fender*

The neural network model based on the YOLO model is tested with the data previously discussed in the data collection section. The precision, the recall, the average precision, and the mean average precision are all metrics used to validate the model. Figure 26 and Table 3 show results of precision, recall, and average precision for each class object.

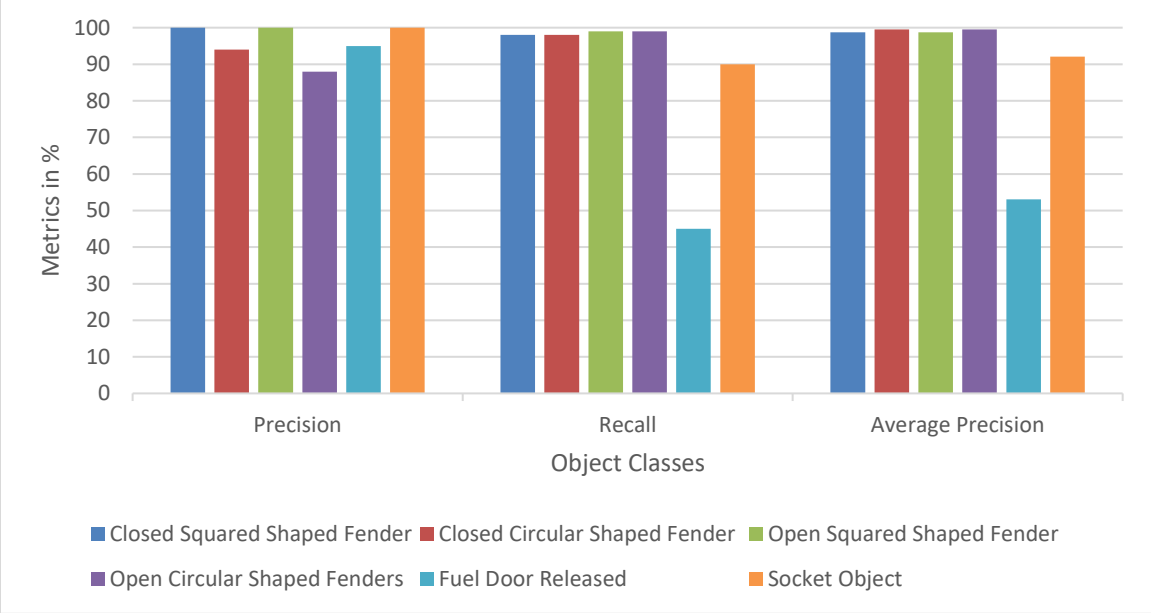


**Figure 26 - Precision, Recall, Average Precision Comparison**

**Table 3 - Results for Object Detection Model**

| | Precision (%) | Recall (%) | Average Precision (%) |
|---|---|---|---|
| Close Squared Shaped Fender | 100 | 98 | 98.7 |
| Close Circular Shaped Fender | 94 | 98 | 99.52 |
| Open Squared Shaped Fender | 100 | 99 | 98.75 |
| Open Circular Shaped Fender | 88 | 99 | 99.56 |
| Fuel Door Released Object | 95 | 45 | 53.08 |
| Socket Object | 100 | 90 | 92.11 |

Based on the results, all class objects contain at least a high precision score of 88 %. The higher precision scores tell that the model correctly predicted most ground truth labels in the validation images. The close and open squared shaped fenders and the socket object had a highest precision score of 100 %, whereas the close and open circular shaped fenders and the fuel door released object have a precision score of 94%, 88%, and 95%, respectively. The model confuses the circular shaped fenders with the fuel released door object. Both class objects have similar circular shapes. However, compared to the squared shaped fenders and the socket objects, the squared shaped fender and the socket object have the distinct shape or the motif. As a result, the model struggled to classify the circular shaped objects and the fuel released door object.

Moving to the recall metrics, most object contains a higher recall score of 90 % except for the fuel door released object had a lowest recall score of 53 %. The higher recall score tells that the model detects all objects on the test images. The fuel door released door object had a lowest recall score due to the small size of the object. During the training period, the data augmentation adds noises such as a blur on the training images. Because of this, the small object often becomes invisible on the images, and the model may not consider the fuel released door as an object class. Additionally, the test sets contain a blurred image with different lighting conditions shown in Figure 27. Depend on the lighting condition, the test images contain a blurred image. Because the fuel released door is a small object relative to other object, the model had frequently failed to identify the fuel released door object with different lighting conditions. Overall, the recall metrics verified the high performance of the model that detects possible targets object on the test images except for the fuel door release object.

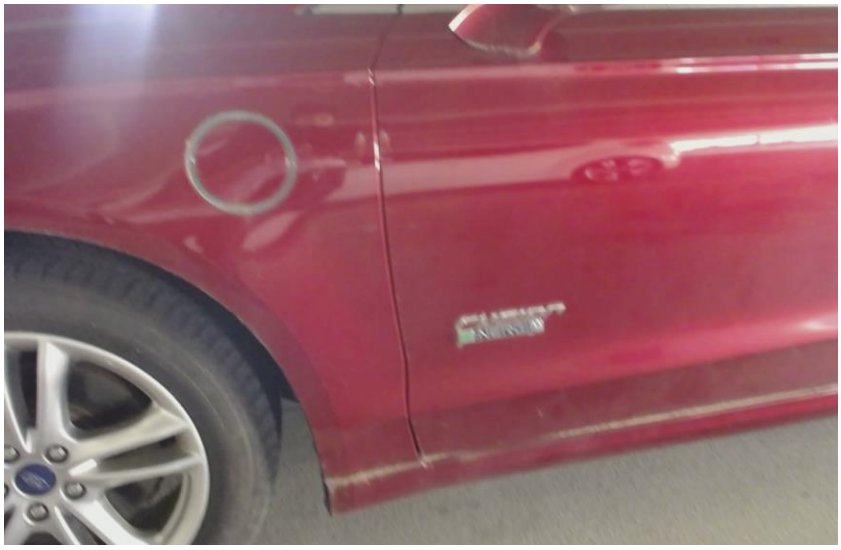

**Figure 27 – An example of the blurred image**

Table 4 shows the overall performance of the CNN model that contains a precision score of 95 % and a recall score of 78 %. The recall score is relatively low due to the low recall score of the fuel door released object as described in the previous section. The mean average precision score shows a high score of 90.29 %. The F-1 score is used to select the best neural network model during the training process. With the custom datasets and the validation dataset, the highest F-1 score is 86 %. Comparing to the target goals, most object classes have a higher precision score of 90 % as well as a higher recall score of 90 % except for the fuel door release object and the open circular shaped fender. Overall, the model performance demonstrated a success in terms of detecting a target object with a high mean average precision of 90.29 %.

**Table 4 – Overall Results**

| Metrics | Results |
|---|---|
| True Positive | 441 |
| False Positive | 21 |
| False Negative | 121 |
| Precision | 95 % |
| Recall | 78 % |
| Mean Average Precision | 90.29 % |
| F-1 Score | 0.86 |

## 4.2 Depth Estimation

### *4.2.1 Evaluation Method*

This section discusses results for the depth estimation models with the ground truth data. The ground truth labels were obtained by measuring the distance between the electric vehicle and the camera attached to the end effector on the robot using a tape measurement. The annotations contain multiple depth measurements from 7 to 20 inch as one set of test data. There are five validation datasets that are used to validate the depth estimation model. The performance of the neural network-based models is compared with the triangle similarity-based depth estimation method. The percent error is used as a primary metric to quantify the performance of each distance detection module shown below.

$$Percent\ Error = \frac{Predicted\ Distance - Ground\ Truth\ Distance}{Ground\ Truth\ Distance} * 100$$

The average percent error is also used to quantify overall results of each model shown in the equation below. The percent error is calculated at each distance measurement. Once all average measurements are computed, the sum of the average error is divided by the total numbers of the distance measurements. The target goal is to obtain the average predicted depth as close as the ground truth labels.

$$Average\ Percent\ Error = \frac{\sum_{q=1}^{Q} Percent\ Error}{Q}$$

$$Q = Total\ Number\ of\ Distance\ Measurements$$

$$q = Percent\ Error\ at\ Each\ Distance\ Measurement$$

In the next section, results from the triangle similarity and the neural network-based distance estimation methods are discussed with the percent error metric.

### *4.2.2 Results*

This section focuses on the results obtained from the triangle similarity and the neural network-based approach. Figure 28 shows that the triangle similarity and the neural network-based approach share their similarity in the predicted distance measurements. Compared to the ground truth labels, both triangle similarity and neural network models estimated the depth smaller than actual depth measurements between 17 cm to 23 cm.

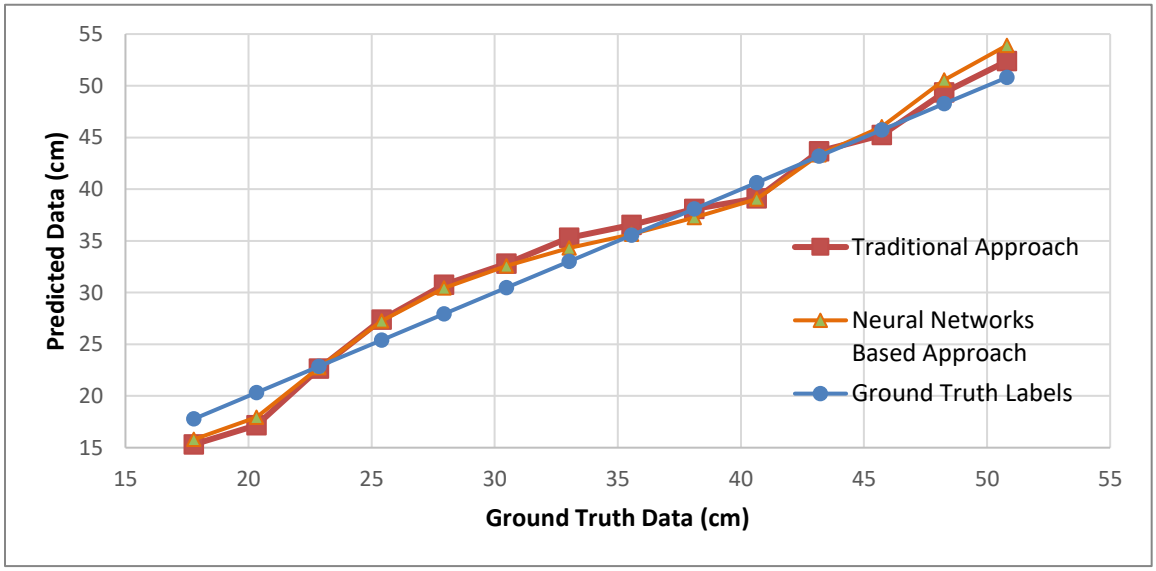


**Figure 28 – Depth Estimation Results**

This result confirmed that both models struggled to predict the depth near the charging points. The size of the bounding box fluctuated more frequently near the charging point. This noise can affect both models because both models depends on the inputs of the bounding box size. As the depth increases from 23 cm to 52 cm, the models estimate the

depth more than the ground truth labels except for the point at 42 cm. The lighting condition may cause such a deviation because the size of the bounding is sensitive to the. The width of the bounding box is sometimes unstable even if the distance measurements are measured at the same location as shown in Figure 29. The size of the bounding box on the left is slightly larger than the actual size of the charging point. Because of the unstable size of the bounding box, both models have errors in their prediction.


**Figure 29 – Examples of Different Lighting Conditions**

Moving to the percent error metrics, Figure 30 and Table 5 show the average estimated distance and percentage errors by each method. The overall percentage error is 5.48 % for the neural network-based method and 4.97 % for the triangle similarity method. Both methods had a large percentage errors at 17.78 cm and 20.32 cm. These results again show that if the camera gets too close to the charging points, both methods had a larger error. The triangle similarity method has smaller percent errors between 17 cm to 35 cm. However, the neural networks approach shows smaller percent errors from 38 cm to 50 cm except for the point at 43 cm and 45 cm. Both models show a trend that the percent error becomes smaller as the depth measurements increase from 17 cm to 50 cm. Because the

near range measurement needs to be precise for inserting the charging connector, this error can be minimized by adding an offset value. Based on this result, the neural network method predicted the depth more accurately than the triangle similarity method.

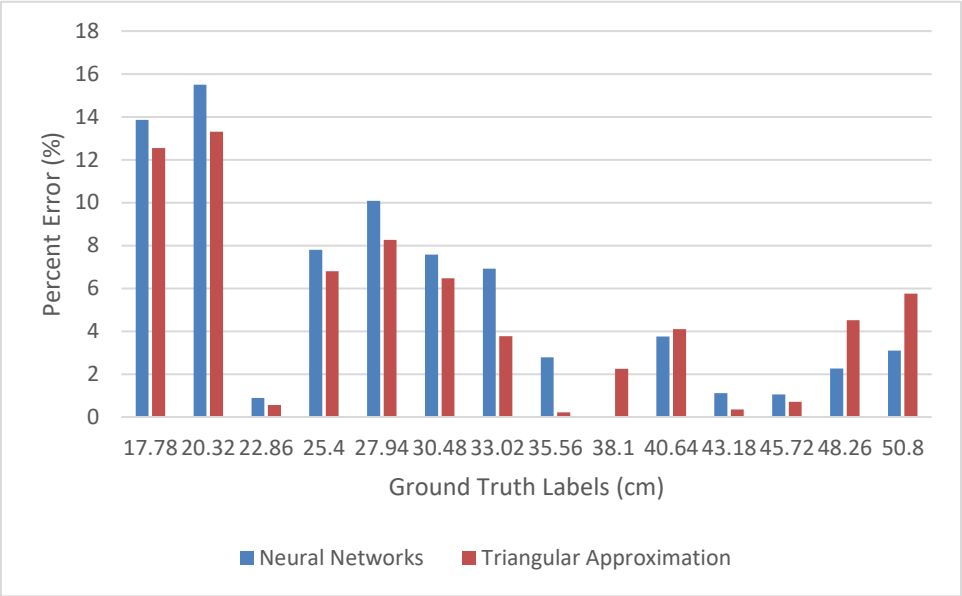


**Figure 30 – Results of the Percent Errors**

**Table 5 – Results for Depth Estimation Tests**

| | NN Model Approach | | Triangle Similarity Approach | |
|---|---|---|---|---|
| Actual Distance (cm) | Average Estimated Distance (cm) | Percentage Error (%) | Average Estimated Distance (cm) | Percentage Error (%) |
| 17.78 | 15.32 | 13.86 | 15.79 | 12.54 |
| 20.32 | 17.17 | 15.50 | 17.9324 | 13.31 |

| | | | | |
|---|---|---|---|---|
| 22.86 | 22.66 | 0.89 | 22.733 | 0.56 |
| 25.4 | 27.38 | 7.80 | 27.2542 | 6.80 |
| 27.94 | 30.76 | 10.09 | 30.4546 | 8.26 |
| 30.48 | 32.79 | 7.58 | 32.5882 | 6.47 |
| 33.02 | 35.31 | 6.92 | 34.3154 | 3.77 |
| 35.56 | 36.55 | 2.79 | 35.6362 | 0.21 |
| 38.1 | 30.8 | 0 | 37.2618 | 2.25 |
| 40.64 | 39.12 | 3.75 | 39.0398 | 4.10 |
| 43.18 | 43.66 | 1.12 | 43.3324 | 0.35 |
| 45.72 | 45.24 | 1.06 | 46.0502 | 0.72 |
| 48.26 | 49.35 | 2.26 | 50.546 | 4.52 |
| 50.8 | 52.37 | 3.10 | 53.8988 | 5.75 |

Regarding the depth data distributions, Figure 31 shows overall depth distributions for the triangle similarity and the neural network approach. Based on the results from Figure 32, the neural network approach generally has lower standard deviation per index distance than the triangle similarity approach. The average standard deviations are 0.59 cm and 0.71 cm for the neural network and triangle similarity approach, respectively. The results confirm that the neural network approach has more consistent predictions per index distance than the triangle similarity approach.

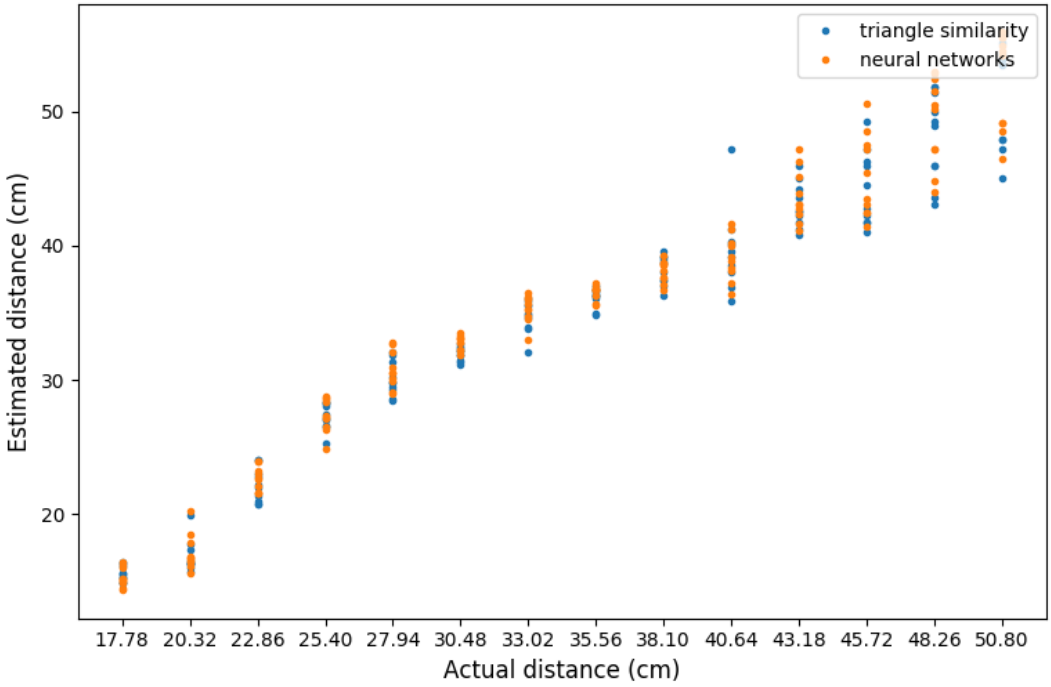


**Figure 31 - Depth Distributions**

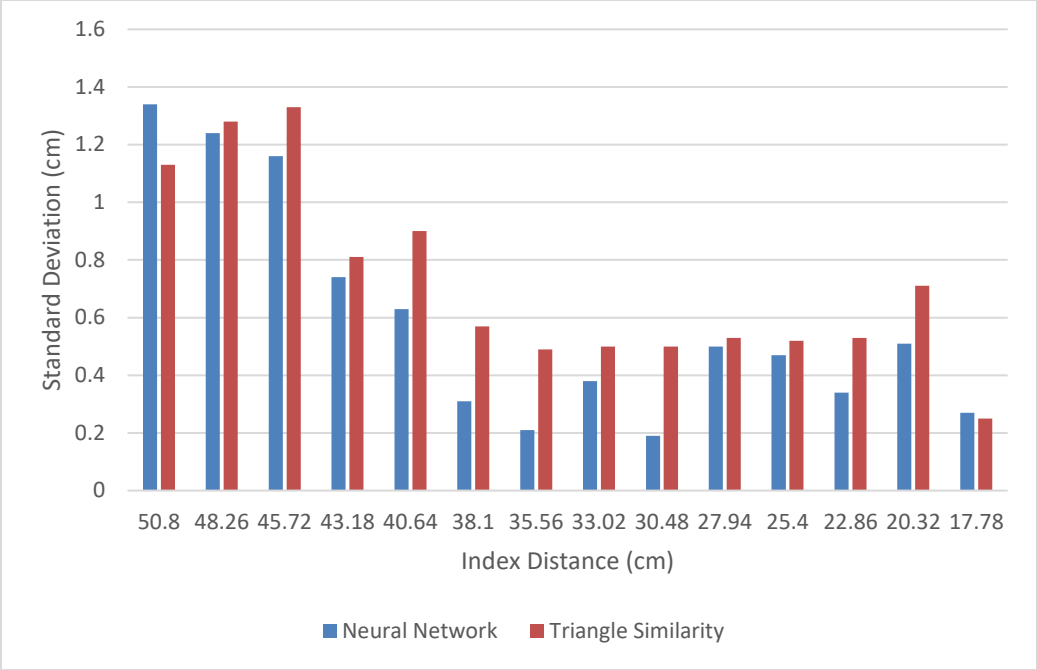


**Figure 32 – Results of Standard Deviations**

Overall, the performance of the neural network-based depth estimation is close to the performance of the triangle similarity approach. Without providing an intrinsic camera parameter and size of target objects, the neural network-based depth detection shows that it can estimate a ground truth depth as close as the triangle similarity approach. Considering the performance of the triangle similarity approach and the target goals, the depth measurements has a relatively small average percent error.

## 4.3 Conclusion

Results from the object detection method using the precision, recall, and mean average precision demonstrates a high performance on localizing and classifying object classes used in the automated charging robot. The result from the neural network-based depth estimation model also shows the low percent error of 5.48 %. Based on the experimental

results, the transfer learning method using the YOLOV3 shows a possible solution as an object detection model. The depth estimation model also provides an insight to estimate the depth on the test images using the information from the object detection model.

# CHAPTER 5. CONCLUSION AND FUTURE WORK

## 5.1 Conclusion

The concept of the object detection and the depth estimation using deep learning techniques are illustrated with the vision system based automated charging station. The object detection using the YOLOV3 architecture enables the accurate charging point detection system in real-time. The neural networks-based depth estimation method shows a potential solution to detect the distance using camera sensors alone. In addition to the autonomous charging station, this thesis demonstrated applications of the camera-based vision system that can apply to autonomous system or mobile robot applications.

### *5.1.1 Object Detection and Recognition*

The vision system uses a neural network trained with the YOLO V3 architecture. The transfer learning method is used here to improve the performance of the convolutional neural networks during the training phase. With the custom dataset, the vision system can detect various targets with different lighting conditions in real time. The vision system model contains a precision and a recall metrics of 95 % and 78 % respectively. The precision score is relatively high enough, but the recall metric is low due to the low recall score of the fuel release door object. Overall, the vision system performs well on the custom dataset.

### *5.1.2 Depth Reconstruction*

The Neural networks-based depth estimation algorithm is used to estimate the distance between the charging point and the end effector on the robot. The triangle similarity approach and the ground truth labels are both used to compare the result of the neural networks-based depth estimation method. The result of the average percentage error metric shows that both methods have similar results that are 5.48 % for the neural network method and 4.97 % for the triangle similarity method. Based on the results, both depth estimation models are sensitive to the noise from the object detection system. Both models rely on the bounding box size that is the output of the object detection system. The neural networks-based depth estimation is preferred over the triangle similarity method in the autonomous charging robot project because the neural network model does not require hand tuning for the camera intrinsic parameters.

## 5.2 Future Work

This thesis covers some of important techniques used in the application of computer vision and deep learning fields. Because this project cooperated with automotive industries, this section includes industry perspectives to improve the automated charging robot. With the rapid development of sensor technology, machine learning, and deep learning field, the vision system can be improved with the following ideas.

Regarding object detection system, the neuro networks can be upgraded to place a 3 – D bounding box over a target in an image to identify the target position and orientation. This can help the end effector to adjust the angle offsets between the robot and the charging point. Additionally, the current neural networks can be improved by providing more

annotations data. With a large dataset, the neural network can observe a various shapes and colors of target objects.

Moving to the depth estimation, using a depth map collected from a stereo vision camera can be used to train a model to estimate the depth between the robot and the charging point. The depth map provides a distance information with respect to the camera sensor. The trained neural network is then implemented on the monocular camera sensor attached to the end effector.

From industry perspective, using a monocular camera can save business cost for company. The lidar and stereo vision sensors are often more expensive than the monocular camera. Additionally, the monocular camera takes less mounting space than the lidar and stereo vision camera. With these potential improvements, the vision system can assist the robot to navigate its end effector to the charging port more accurately in real time.

# APPENDIX

## A. Appendix A

All data collected from the experiments are shown in Appendix A. Table 6 through Table 11 contains additional information about validating the result from the object detection model.

**Table 6 – Result for Closed Squared Shaped Fender Class**

| Metrics | Result |
|---|---|
| True Positive | 78 objects |
| False Positive | 0 |
| Precision | 100 % |
| Recall | 98 % |
| Average Precision | 98.70 % |

**Table 7 – Result for Closed Circular Shaped Fender Class**

| Metrics | Result |
|---|---|
| True Positive | 78 |
| False Positive | 5 |
| Precision | 94 % |
| Recall | 98 % |
| Average Precision | 99.52 % |

**Table 8 - Results for Open Squared Shaped Fender Class**

| Metrics | Result |
|---|---|
| True Positive | 79 |
| False Positive | 0 |
| Precision | 100 % |
| Recall | 99 % |
| Average Precision | 98.75 % |

**Table 9 - Results for Open Circular Shaped Fenders**

| Metrics | Result |
|---|---|
| True Positive | 79 |
| False Positive | 11 |
| Precision | 88 % |
| Recall | 99 % |
| Average Precision | 99.56 % |

**Table 10 – Results for Fuel Door Released Object**

| Metrics | Results |
|---|---|
| True Positive | 92 |
| False Positive | 5 |
| Precision | 95 % |
| Recall | 45 % |
| Average Precision | 53.08 % |

**Table 11 – Results for Socket Object**

| Metrics | Results |
|---|---|
| True Positive | 35 |
| False Positive | 0 |
| Precision | 100 % |
| Recall | 90 % |
| Average Precision | 92.11 % |

## B. Appendix B

This section provides additional images used in the training and validation sets for the object detection system.

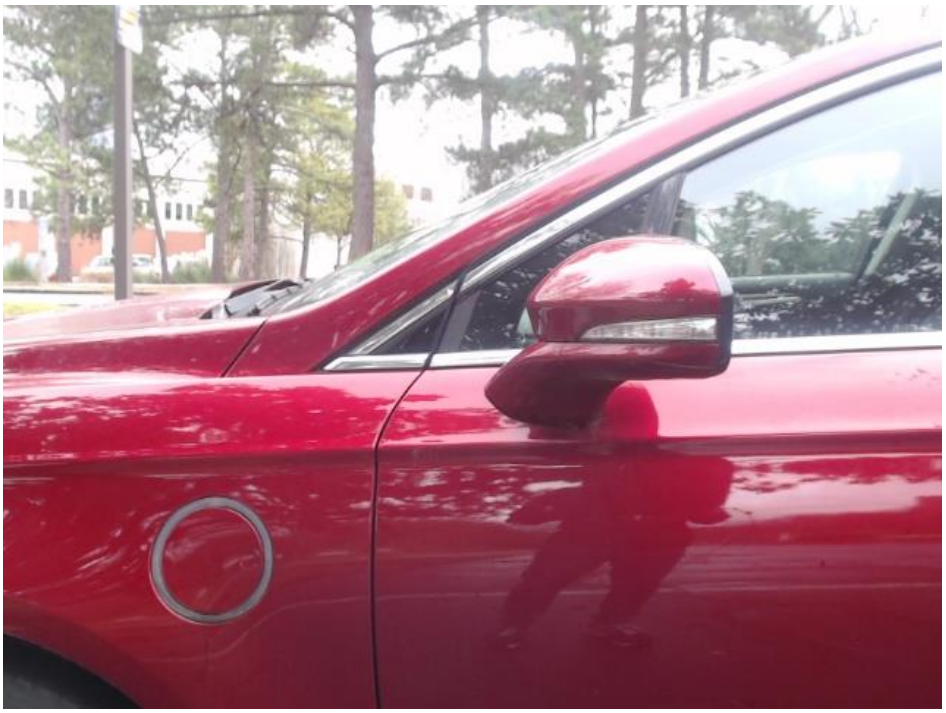

**Figure 33 – An example of the closed circular red fender in the outdoor environment**

**Figure 34 – An example of the closed circular red fender in the indoor environment**

**Figure 35 – An example of the closed circular red fender in the indoor environment**

**Figure 36 – An example of the closed circular blue fender int he indoor environment**

**Figure 37 – An example of the closed circular blue fender in the outdoor environment**

**Figure 38 – An example of the closed champagne fender in the outdoor environment**

**Figure 39 – An example of the opened champagne fender in the outdoor environment**

# REFERENCES


1. Brown, B. *Evidence stacks up in favor of self-driving cars in 2016 NHTSA fatality report*. 2017; Available from: https://www.digitaltrends.com/cars/2016-nhtsa-fatality-report/.

2. ENERGY, O.o.E.E.R. *FOTW #1057, November 26, 2018: One Million Plug-in Vehicles Have Been Sold in the United States*. 2018 November 26; Available from: https://www.energy.gov/eere/vehicles/articles/fotw-1057-november-26-2018-one-million-plug-vehicles-have-been-sold-united.

3. Kane, M., *U.S. Plug-In Car Sales (cummulative)*. 2018, INSIDEEVs.

4. McCarthy, N., *The Evolution Of U.S. Electric Vehicle Charging Points.* 2018.

5. Intelligence, B.I. *10 million self-driving cars will be on the road by 2020*. 2016; Available from: https://www.businessinsider.com/report-10-million-self-driving-cars-will-be-on-the-road-by-2020-2015-5-6.

6. Cha, J. *Hyundai Motor Group Unveils Innovative Electric Vehicle Charging and Automated Parking Systems Concept*. 2019; Available from: https://www.hyundainews.com/en-us/releases/2678.

7. energysage. *Range of electric vehicles*. 2019; Available from: https://www.energysage.com/electric-vehicles/buyers-guide/range-of-distance-for-top-evs/.

8. Yajima, Y., *Jetson TX2*. 2018, Yosuke Yajima: Georgia Institute of Technology.

9. Ogura, T. *cv_camera*. 2019; Available from: http://wiki.ros.org/cv_camera.

10. Joseph Redmon, A.F., *YOLOv3: An Incremental Improvement.* 2018.

11. Alex. *Yolo-v3 and Yolo-v2 for Windows and Linux*. 2016 [cited 2020 4/13]; Available from: https://github.com/AlexeyAB/darknet.

12. Bjelonic, M. *YOLO ROS: Real-Time Object Detection for ROS*. 2019; Available from: https://github.com/leggedrobotics/darknet_ros.

13. Tzutalin. *labelImg*. 2015; Available from: https://github.com/tzutalin/labelImg.

14. Yajima, Y., *Automated Charging Station*. 2018, Yosuke Yajima: Georgia Institute of Technology.

15. Mark. *Calibration Checkerboard Collection*. 2018; Available from: https://markhedleyjones.com/projects/calibration-checkerboard-collection.

16. James Bowman, P.M. *camera_calibration*. 2019; Available from: http://wiki.ros.org/camera_calibration.

17. Sharma, K. and N. Thakur, *A review and an approach for object detection in images.* International Journal of Computational Vision and Robotics, 2017. **7**: p. 196.

18. Singh Chahal, K. and K. Dey *A Survey of Modern Object Detection Literature using Deep Learning*. arXiv e-prints, 2018.

19. Ali, M.H., et al., *Vision-based Robot Manipulator for Industrial Applications.* Procedia Computer Science, 2018. **133**: p. 205-212.

20. Han, S., *ADVANCED DRIVER ASSISTANCE SYSTEMS TECHNIQUES USING A SINGLE CAMERA AND LIDAR*, in *School of Electrical and Computer Engineering*. 2018, Georgia Institute of Technology: Georgia Institute of Technology. p. 44.

21. Sridhar, S., *COMPUTER VISION FOR DRIVER ASSISTANCE SYSTEMS*, in *School of Electrical and Computer Engineering*. 2018, Georgia Institute of Technology: Georgia Institute of Technology.

22. Bahadur, S., et al., *LITERATURE REVIEW ON VARIOUS DEPTH ESTIMATION METHODS FOR AN IMAGE.* LITERATURE REVIEW ON VARIOUS DEPTH ESTIMATION METHODS FOR AN IMAGE, 2017. **VOL 5**: p. 8-13.

23. Jamwal, N., N. Jindal, and K. Singh. *A survey on depth map estimation strategies*. in *International Conference on Signal Processing (ICSP 2016)*. 2016.

24. Al-Jarrah, M.A. *Developing 3D model for mobile robot environment using mono-vision system*. in *2016 7th International Conference on Computer Science and Information Technology (CSIT)*. 2016.

25. Saxena, A., S.H. Chung, and A.Y. Ng, *3-D Depth Reconstruction from a Single Still Image.* International Journal of Computer Vision, 2008. **76**(1): p. 53-69.

26. Electrify America, L. *Electrify America And Stable Announce Collaboration to Deploy Robotic Fast-Charging Facility for Self-Driving Electric Vehicle Fleets*. August 1, 2019; Available from: https://www.prnewswire.com/news-releases/electrify-america-and-stable-announce-collaboration-to-deploy-robotic-fast-charging-facility-for-self-driving-electric-vehicle-fleets-300894811.html.

27. Yvkoff, L., *Robotic EV Fast-Charging Stations Planned For San Francisco*, S.i.c.w.E.A.t.t.r.E.f.-c.s.i.S. Francisco., Editor. 2019, Forbes.

28. Fronzek, T. *e-smartConnect: Volkswagen is conducting research on an automated quick-charging system for the next generation of electric vehicles*. July 13, 2015; Available from: https://www.volkswagen-newsroom.com/en/press-

releases/e-smartconnect-volkswagen-is-conducting-research-on-an-automated-quick-charging-system-for-the-next-generation-of-electric-vehicles-1512.

29. *Roadmap E: full of energy!* 2017, Volkswagen AG 2018.

30. Lambert, F., *Here's a real robot electric car fast-charging station like the one Tesla has been promising for years*. 2018, electrek.

31. KOOSER, A. *Tesla's robo-snake charger prototype is our new car overlord*. AUGUST 6, 2015; Available from: https://www.cnet.com/roadshow/news/teslas-robo-snake-charger-prototype-is-our-new-car-overlord/.

32. Lambert, F., *Tesla patent shows new way to automated high-speed charging with external cooling*. April 22, 2017, electrek.

33. VOLTERIO. *FULLY AUTOMATIC CHARGING*. [cited 2019 October 10]; Available from: http://www.volterio.com/index.html.

34. VOLTERIO, *FULLY AUTOMATIC CHARGING*. VOLTERIO.

35. Redmon, J., et al., *You Only Look Once: Unified, Real-Time Object Detection*, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, T. Model, Editor. 2016. p. 779-788.

36. Joseph Redmon, A.F., *Darknet-53*, Darknet-53, Editor. 2018: Cornell University.

37. Han, S., *Coordinate System Conversion from 2-D to 3-D using a Pinhole Camera Model*. 2018, Georgia Institute of Technology: Georgia Instutute of Technology.

38. HOLLEMANS, M.I., *One-stage object detection*, H.w.d.t.m. work?, Editor. 2018.

39. ENERGY, O.o.E.E.R. *FOTW #1089, July 8, 2019: There are More Than 68,800 Electric Vehicle Charging Units in the United States*. 2019 July 8; Available from: https://www.energy.gov/eere/vehicles/articles/fotw-1089-july-8-2019-there-are-more-68800-electric-vehicle-charging-units.

40. Fiala, M. *Vision guided control of multiple robots*. in *First Canadian Conference on Computer and Robot Vision, 2004. Proceedings.* 2004.